

From THE INSTITUTE OF ENVIRONMENTAL MEDICINE,  
DEPARTMENT OF CARDIOVASCULAR EPIDEMIOLOGY  
Karolinska Institutet, Stockholm, Sweden

# **UTILITIES AND QUALITY OF LIFE IN CARDIOVASCULAR DISEASE – DRIVERS, ECONOMIC EFFECTS AND CLINICAL OUTCOMES**

Jenny Berg



**Karolinska  
Institutet**

Stockholm 2014

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Åtta.45 Tryckeri AB, Solna, Sweden

© Jenny Berg, 2014

ISBN 978-91-7549-716-7

# Utilities and quality of life in cardiovascular disease – drivers, economic effects and clinical outcomes THESIS FOR DOCTORAL DEGREE (PhD)

By

**Jenny Berg**

*Principal Supervisor:*

Peter Lindgren, PhD  
Karolinska Institutet  
Department of Learning, Informatics,  
Management and Ethics  
Medical Management Centre

*Co-supervisors:*

Professor Thomas Kahan  
Karolinska Institutet  
Department of Clinical Sciences, Danderyd Hospital  
Division of Cardiovascular Medicine

Carolin Miltenburger, PhD  
Novartis Pharma AG  
Global Health Economics and Outcomes  
Research

*Opponent:*

Professor Jan Sørensen  
University of Southern Denmark  
Institute of Public Health  
Centre for Applied Health Services Research, Centre for  
Health Economic Research

*Examination Board:*

Professor Lars-Åke Levin  
Linköping University  
Department of Medical and Health Sciences  
Centre for Medical Technology Assessment

Maria Schaufelberger, MD, PhD  
University of Gothenburg  
Sahlgrenska Academy  
Department of Molecular and Clinical Medicine

Niklas Zethraeus, PhD  
Karolinska Institutet  
Department of Learning, Informatics,  
Management and Ethics  
Medical Management Centre



*To my family*

# ABSTRACT

Atrial fibrillation and chronic heart failure represent important contributors to cardiovascular disease burden, with high incidence, prevalence and mortality rates. The overall objective of this thesis is to enhance the knowledge on what drives preferences for different health states (utilities) in patients with atrial fibrillation and chronic heart failure, and how health-related quality of life and utilities in turn influence clinical and economic outcomes.

In Paper I, we used data from a European-wide observational study of patients with atrial fibrillation treated in cardiology clinics to estimate determinants of utility based on the EQ-5D. At baseline, increasing age, female gender, domestic status outside the own home, existing comorbidities, and symptoms of atrial fibrillation, chronic heart failure or angina were associated with reduced utility, while regular physical activity had a positive effect. At 1-year follow-up, significant determinants included atrial fibrillation symptoms and major adverse events, including stroke, myocardial infarction and chronic heart failure.

In Paper II, we applied some of the results from Paper I in an economic evaluation of the anti-arrhythmic treatment dronedarone based on patient-level data from the ATHENA trial. The within-trial analysis indicated that dronedarone when used as in ATHENA is cost-effective within generally accepted thresholds (base case incremental cost-effectiveness ratio CAD\$7560 per quality-adjusted life year), and that this also would hold in the light of subsequent label restrictions.

In Paper III, we used data from the Swedish national registry on chronic heart failure, which mainly covers hospitalised patients, to analyse drivers of utility based on the EQ-5D. Utility at baseline was negatively affected by female gender, increasing age, increasing New York Heart Association class, preserved left ventricular ejection fraction, lung disease, diabetes, and use of nitrates, antiplatelets or diuretics. Higher systolic blood pressure and haemoglobin levels and use of angiotensin converting enzyme inhibitors/angiotensin receptor blockers or beta-blockers were associated with increased utility. A significant interaction between age category and functional class indicated that patients in the youngest age group are more severely affected by worsening functional status than older patients. In our data set, the ordinary least squares model performed slightly better than the two-part model on a population level and for capturing utility ranges.

In Paper IV, we investigated the role of health-related quality of life measured with a generic instrument on clinical outcomes in patients hospitalised with systolic heart failure as part of a randomised controlled study in Sweden. Physical mobility was a significant independent predictor for all-cause and cardiovascular rehospitalisation and mortality, with every 1% worsening resulting in a 1-2% increase in the hazard ratio of being hospitalised or dying. Emotional reactions were an additional independent predictor for all-cause hospitalisations, with a similar impact as physical mobility. Additional analyses suggest that the impact of health-related quality of life, specifically physical mobility, on cardiovascular mortality may be similar regardless of timing and setting of assessment.

In summary, the studies in this thesis support the use of health-related quality of life and utilities as a value-added part of clinical and economic decision-making, due to their relationship with both clinical and economic outcomes.

## LIST OF SCIENTIFIC PAPERS

- I. Berg J, Lindgren P, Nieuwlaat R, Bouin O & Crijns H (2010). Factors determining utility measured with the EQ-5D in patients with atrial fibrillation. *Qual Life Res*; 19: 381-390.
- II. Berg J, Sauriol L, Connolly S & Lindgren P (2013). Cost-effectiveness of dronedarone in patients with atrial fibrillation in the ATHENA trial. *Can J Cardiol*; 29: 1249-1255.
- III. Berg J, Lindgren P, Mejhert M, Edner M, Dahlström U & Kahan T. Determinants of utility based on the EQ-5D in chronic heart failure patients and their change over time: results from the Swedish Heart Failure Registry. Submitted.
- IV. Berg J, Lindgren P, Kahan T, Schill O, Persson H, Edner M & Mejhert M (2014). Health-related quality of life and long-term morbidity and mortality in patients hospitalised with systolic heart failure. *JRSM Cardiovasc Dis*; 3: 2048004014548735.

# CONTENTS

1	Introduction .....	1
2	Background.....	3
2.1	Cardiovascular disease burden.....	3
2.1.1	Atrial fibrillation .....	3
2.1.2	Heart failure.....	5
2.2	Methodological concepts.....	7
2.2.1	Health-related quality of life and utilities.....	7
2.2.2	Economic evaluations .....	15
2.2.3	Risk stratification for prognosis and management.....	21
2.3	Review of the literature .....	23
2.3.1	Health-related quality of life and utilities in atrial fibrillation.....	23
2.3.2	Economic evaluations in atrial fibrillation and the impact of utilities .....	26
2.3.3	Health-related quality of life and utilities in chronic heart failure.....	29
2.3.4	Impact of health-related quality of life on clinical outcomes in chronic heart failure .....	33
2.3.5	Identified knowledge gaps .....	36
3	Aims of the thesis .....	37
4	Materials and methods .....	39
4.1	Materials .....	39
4.1.1	The Euro Heart Survey on Atrial Fibrillation .....	39
4.1.2	The ATHENA trial.....	40
4.1.3	The Swedish Heart Failure Registry.....	41
4.1.4	The OPTIMAL study .....	42
4.2	Methods .....	43
4.2.1	Regression analysis .....	43
4.2.2	Cost-effectiveness analysis .....	45
5	Results.....	47
5.1	Determinants of utility in atrial fibrillation (Paper I) .....	47
5.2	Role of different outcome measures in cost-effectiveness analyses of a new treatment for atrial fibrillation (Paper II) .....	48
5.3	Determinants of utility in chronic heart failure (Paper III) .....	49
5.4	Impact of health-related quality of life on prognosis in chronic heart failure (Paper IV).....	50
5.4.1	Previously unpublished results: mortality impact of health-related quality of life measured at 1-year follow-up visit in outpatient care.....	52
6	Discussion.....	53
6.1	Critical evaluation of study methods and findings .....	53
6.1.1	Representativeness of study samples.....	54
6.1.2	Meaningfulness of health-related quality of life measurements.....	55
6.1.3	Considerations when using utilities from different studies.....	57



6.2	The role of health-related quality of life and utilities in public health.....	58
6.2.1	Health policy applications: use by payers .....	58
6.2.2	Prognostic markers in clinical practice.....	61
6.2.3	Possible use with patients .....	61
6.3	Suggestions for future research.....	62
7	Conclusions .....	65
8	Acknowledgements .....	67
9	References .....	69

## LIST OF ABBREVIATIONS

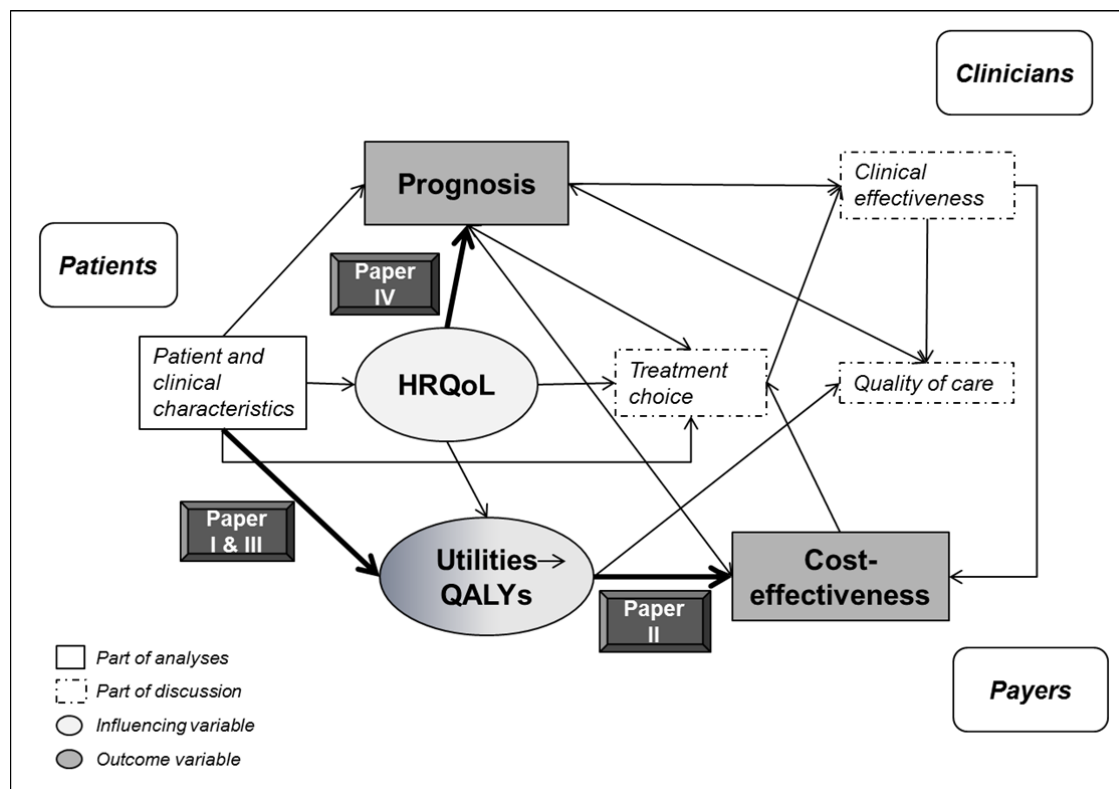
AF	Atrial fibrillation
CHF	Chronic heart failure
CLAD	Censored least absolute deviations
DALY	Disability-adjusted life year
EQ-5D	EuroQoL-5D
HFPEF	Heart failure with preserved ejection fraction
HFREF	Heart failure with reduced ejection fraction
HRQoL	Health-related quality of life
HTA	Health technology assessment
HUI	Health Utilities Index
ICER	Incremental cost-effectiveness ratio
IQWiG	Institute for Quality and Efficiency in Health Care
KCCQ	Kansas City Cardiomyopathy Questionnaire
LVEF	Left ventricular ejection fraction
MLWHFQ	Minnesota Living With Heart Failure Questionnaire
NHP	Nottingham Health Profile
NICE	National Institute for Health and Care Excellence
NYHA	New York Heart Association
OLS	Ordinary least squares
QALY	Quality-adjusted life year
SF-36	Short Form-36
TLV	Swedish Dental and Pharmaceutical Benefits Agency
WHO	World Health Organization
WTP	Willingness to pay

# 1 INTRODUCTION

When I started working on this thesis, I had been looking for a topic that provided a combination of interesting theory, new developments and applications to different areas of health economics and health care. I have always found the area of quality of life highly fascinating, as it combines many interesting aspects, with its origins in philosophy and its applications to the fields of medicine, economics and health care, which also have given rise to many methodological approaches and issues. During the thesis, I have come to better understand the complexities and challenges in this field and how it ties into many areas within public health. Every thesis is of course a learning experience, and in this regard there have been some interesting methodological developments in the field of how quality of life weights, or utilities, should be analysed, which has informed the included studies on what drives utility in two important cardiovascular disorders.

Figure 1 presents an attempt at schematically summarising the different areas covered by the thesis and some of their relationships also with other relevant topics in the field. The central theme is quality of life (measured and analysed in different ways), and the included studies seek to investigate what factors affect it in patients with selected cardiovascular diseases, and what role it plays in various areas of health care and public health.

**Figure 1: Conceptual positioning of the thesis, including areas and relationships covered by the different studies**



Notes: HRQOL=health-related quality of life, QALY=quality-adjusted life year



## 2 BACKGROUND

### 2.1 CARDIOVASCULAR DISEASE BURDEN

Cardiovascular diseases, covering diseases of the heart and blood vessels, remain the leading mortality cause globally, accounting for over 17 million deaths in 2008. Within this group, ischaemic heart disease and cerebrovascular diseases are the most common causes of death. Since 1990, mortality from cardiovascular diseases has been reduced in high-income countries as a result of primary and secondary prevention and improved treatment. At the same time, there has been a rapid mortality increase in low- and middle-income countries due to resource shortages and organisational challenges (World Health Organization, 2011).

In Europe, the role of cardiovascular diseases as a cause of mortality is also pronounced, accounting for almost one in two deaths based on latest available years (Nichols et al., European Society of Cardiology, 2012). While mortality is declining in most European countries, disease burden remains high. In 2009, the total societal cost of cardiovascular diseases in the 27 member states of the European Union was estimated at €196 billion, with direct health care costs accounting for 54% of total costs, productivity losses for 24% and informal care for 22% (Leal, Luengo-Fernandez & Gray, European Society of Cardiology, 2012).

Apart from significant mortality and high economic costs, cardiovascular disease also leads to substantial reductions in health and quality of life. The World Health Organization (WHO) uses the disability-adjusted life year (DALY) to capture both time losses due to early death and time lived with reduced health compared to the norm (World Health Organization, 2013). In 2011, ischaemic heart disease and stroke led to 295 million DALYs lost and were amongst the top three causes of DALYs lost globally (World Health Organization, 2013).

The focus in this thesis is on two cardiovascular indications that are significant due to the comparatively high risk for subsequent morbidity and mortality, making them important contributors to global cardiovascular disease burden: atrial fibrillation (AF) and chronic heart failure (CHF).

#### 2.1.1 Atrial fibrillation

##### 2.1.1.1 *Definition and diagnosis*

AF is a cardiac arrhythmia characterised by specific electrocardiographic patterns (see e.g. Camm et al., 2010). Common symptoms include heart palpitations, shortness of breath, fatigue and weakness. The European Society of Cardiology groups AF into five types (Camm et al., 2010):

- First diagnosed AF: applies to all patients who present with AF for the first time;
- Paroxysmal AF: arrhythmia stopping by itself, mostly within 48 hours;

- Persistent AF: an episode lasting longer than seven days or requiring pharmacological or electrical cardioversion;
- Long-standing persistent AF: AF lasting at least one year when a rhythm control strategy has been applied;
- Permanent AF: arrhythmia that has been accepted by the patient and physician, meaning that rhythm control interventions are no longer pursued.

In addition, AF can be asymptomatic (silent); in this case it may be diagnosed as part of an electrocardiogram for other reasons or may become apparent in the form of AF-related complications, such as stroke. Another way of classifying AF is into valvular and non-valvular AF, i.e. depending on whether there is an underlying valvular pathology or not.

#### *2.1.1.2 Epidemiology and risk factors*

AF has been termed a growing global epidemic (Chugh et al., 2014; Lip, Kakar & Watson, 2007). Global prevalence estimates based on a recent systematic literature review (Chugh et al., 2014) were 474 per 100 000 persons in 2010, with North America standing out as the region with the highest prevalence (702 per 100 000). According to this review, incidence rates have increased by over 25% from 1990 to 2010, to 77.5 per 100 000 person-years in men and 59.5 in women. At the same time, mortality has almost doubled due to the increases in prevalence and incidence. Since AF is an important risk factor for stroke and heart failure (Ball et al., 2013; Wolf, Abbott & Kannel, 1991), these complications are also set to grow with the increasing incidence of AF.

Two major population-based studies in the US (Framingham) and Europe (Rotterdam, The Netherlands) have independently shown a lifetime risk of around 25% for developing AF from the age of 40 years (Heeringa et al., 2006; Lloyd-Jones et al., 2004). The risk factors for development of AF are manifold. They include increasing age, male gender, genetic predisposition, and a range of cardiovascular, metabolic and inflammatory factors, such as hypertension, diabetes, valvular heart disease, myocardial infarction, heart failure, obesity and raised inflammatory markers (Kannel et al., 1998; Rienstra, McManus & Benjamin, 2012). The rise of some of these risk factors, in particular an ageing population, increased obesity and hypertension, as well as improved survival with other diseases and a better diagnosis, are likely contributors to the rising AF incidence (Chugh et al., 2014). Similarly, as patients survive longer and often have other comorbid cardiovascular conditions, this may explain some of the increase in mortality observed especially in high-income countries.

#### *2.1.1.3 Management*

AF management consists of two parallel therapeutic goals: reduction of symptoms and prevention of severe AF complications. A cornerstone of AF therapy lies in stroke prevention with antithrombotic agents, including aspirin, warfarin or other oral anticoagulants. To reduce the risk of complications, it is also critical to treat concomitant cardiac diseases, such as hypertension and dyslipidaemia. Rate control is usually required to control symptoms.

Additional symptom relief may be needed in the form of rhythm control therapy using cardioversion, antiarrhythmic drugs or ablation. Recent updates of e.g. the European AF guidelines are reflective of several pharmacological developments in this field over the past decade, including the introduction of dronedarone and subsequent restrictions in its indication due to further study results, and the emergence of novel oral anticoagulants with improved safety profiles (Camm et al., 2010; Camm et al., 2012).

Over the last decade, risk stratification for stroke and thromboembolism in AF has moved from the simpler CHADS<sub>2</sub> index (which assigns one point for cardiac failure, hypertension, age and diabetes, and two points for stroke) to the more comprehensive CHA<sub>2</sub>DS<sub>2</sub>-VASc score (which covers congestive heart failure, hypertension, age  $\geq 75$  years [doubled], diabetes, stroke [doubled], vascular disease, age 65-74, and sex category [female]). As antithrombotic therapy increases the risk of bleeding, it is recommended to assess this by using e.g. the HAS-BLED score (hypertension, abnormal renal/liver function, stroke, bleeding history or disposition, labile international normalised ratio, elderly [ $>65$  years], drugs/alcohol concomitantly) (Camm et al., 2010; Camm et al., 2012). Despite these recommendations, large international surveys have shown that clinical guidelines are not followed in terms of indicated antiarrhythmic and anticoagulation therapy (Alam et al., 2012; Nieuwlaat et al., 2005).

## **2.1.2 Heart failure**

### *2.1.2.1 Definition and diagnosis*

HF is a complex syndrome arising from an abnormal functioning or structure of the heart, hindering the heart to fill with or pump a sufficient amount of blood through the body. As several of its typical symptoms are relatively non-specific, it can be difficult to diagnose, and different diagnostic criteria exist (Roger, 2013). Key diagnostic tools include electrocardiogram, echocardiogram and laboratory tests. The European Society of Cardiology has set up the following criteria for the diagnosis of HF (McMurray et al., 2012):

- HF with reduced ejection fraction (HFREF), traditionally also termed systolic HF:
  - Symptoms typical of HF (e.g. breathlessness, fatigue or ankle swelling);
  - Signs typical of HF (e.g. elevated jugular pressure or pulmonary crackles);
  - Reduced left ventricular rejection fraction (LVEF).
- HF with preserved ejection fraction (HRPEF), traditionally also called diastolic HF:
  - Symptoms typical of HF;
  - Signs typical of HF;
  - Normal ( $\geq 50\%$ ) or only mildly reduced LVEF and left ventricle not dilated;
  - Relevant structural heart disease and/or diastolic dysfunction.

In terms of LVEF, there is a “grey zone” where definitions for HFREF and HFPEF may differ across sources. For example, the European Society of Cardiology 2012 guidelines

consider LVEF between 35-50% to mostly represent mild systolic dysfunction (McMurray et al., 2012).

The New York Heart Association (NYHA) classification is a separate, complementary grading system that focuses on severity of symptoms and physical functioning (McMurray et al., 2012):

- Class I: No limitation of physical activity. Ordinary physical activity does not cause undue breathlessness, fatigue or palpitations;
- Class II: Slight limitation of physical activity. Comfortable at rest, but ordinary physical activity results in undue breathlessness, fatigue or palpitations;
- Class III: Marked limitation of physical activity. Comfortable at rest, but less than ordinary physical activity results in undue breathlessness, fatigue or palpitations;
- Class IV: Unable to carry on any physical activity without discomfort. Symptoms at rest can be present. If any physical activity is undertaken, discomfort is increased.

There are many terms to describe different stages or presentations of HF. In this thesis, the focus will lie on chronic HF, i.e. patients who have had the disease for some time. It should be noted that patients with stable CHF may still experience acute (decompensation) phases.

#### *2.1.2.2 Epidemiology and risk factors*

There are a sizable number of epidemiological studies in HF that are based on different populations in Europe and the US, including renowned cohorts from Rotterdam (The Netherlands), Hillingdon (UK), Framingham (US), and Olmsted County (US). Given the different diagnostic criteria used, figures vary across studies, but they all highlight the significant disease burden and important areas of unmet need. The prevalence of CHF is estimated at 1-2% in the Western world and rises sharply at 75 years of age, with a prevalence of more than 10% in those aged over 85 years (Mosterd & Hoes, 2007; Zarrinkoub et al., 2013). Incidence rates vary from 1.3 per 1 000 person-years in populations over 25 years (Cowie et al., 1999) to 14.4 per 1 000 person-years in those aged 55 years or above (Bleumink et al., 2004).

Prognosis remains poor despite some advances in patient management; recently, median survival in hospitalised patients was reported as 1.8 years in women and 2.3 in men, and the 1-year hospital readmission rate as around 30% (Jhund et al., 2009). While a decrease in incidence and mortality rates has been observed for several age groups (e.g. Barasa et al., 2014), the disease burden remains high due to a continued and growing need for specialised and hospital care (Roger, 2013; Zarrinkoub et al., 2013).

The lifetime risk for developing HF in a person aged 55 years has been reported as between 20-30% (Bleumink et al., 2004; Lloyd-Jones et al., 2002). Risk factors for HF include myocardial infarction, hypertension, obesity, valvular heart disease, genetic predisposition, renal dysfunction and obstructive pulmonary disease. AF has also been shown to be



correlated with long-term development of CHF (Ball et al., 2013). In Western countries, the major causes for HF are ischaemic heart disease and hypertension (Mosterd & Hoes, 2007).

### *2.1.2.3 Management*

The therapeutic options for CHF management have been relatively unchanged for the last few decades. More recently, new pharmacological treatment has emerged with e.g. ivabradine, and research activity for previously underserved patient groups, including HFPEF, has shown some promising clinical trial results. Recent reports from a large European registry indicate that treatment guidelines are largely followed, in particular in areas with clear evidence-based recommendations and relatively established treatment experience, such as pharmacological treatment of CHF (Maggioni et al., 2013).

For patients with HFREF, disease-modifying pharmacological treatments exist in the form of angiotensin converting enzyme inhibitors, angiotensin receptor blockers, beta-blockers and mineralocorticoid receptor antagonists. In addition, symptomatic treatment with diuretics is common. In addition, device therapy, such as implantable cardioverter-defibrillators and cardiac resynchronisation therapy, may be useful in certain patient groups. In contrast, for HFPEF, no pharmacological therapy is available to date that has demonstrated any impact on morbidity or mortality. Symptomatic treatment with diuretics and management of comorbidities, such as hypertension, myocardial ischemia and AF, are recommended (McMurray et al., 2012).

## **2.2 METHODOLOGICAL CONCEPTS**

### **2.2.1 Health-related quality of life and utilities**

In this section, terms that are important to the concepts addressed in this thesis are defined and placed into their wider theoretical context. As these concepts partly originate from different disciplines and have been used in different ways even within these, some of the terms do not necessarily have standard definitions. For the purposes of this thesis, the most relevant and appropriate definitions given the addressed research questions are provided.

#### *2.2.1.1 Quality of life*

Quality of life can be an elusive concept that by nature is highly subjective and therefore rather difficult to quantify. In this context, it is worth citing WHO's definition of health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (Preamble to the Constitution of the World Health Organization, Geneva, 1948).

The term health-related quality of life (HRQoL) is used to distinguish between quality of life in a wider sense and the more specific health care and clinical setting. Generally, HRQoL is seen as multidimensional concept, covering several aspects (domains). This is an important difference compared to single-item health status questions, which nevertheless may have a

broader scope than HRQoL in that they refer to all components of health as defined by the WHO.

Several theoretical HRQoL models have been proposed to illustrate the relationships between different related aspects. The most frequently used model was proposed by Wilson and Cleary in 1995 (Wilson & Cleary, 1995); it conceptualises the relationships between overall quality of life and different measures of health, covering biological and physiological variables, symptom status, functional status and general health perceptions.

#### *2.2.1.2 Health-related quality of life instruments*

Given its subjective nature, HRQoL should ideally be measured by asking patients or the individuals affected. A multitude of instruments exist for this, which can cover different domains of HRQoL, such as general health, physical functioning, social functioning, emotional or mental well-being, but also existential questions. Two major types of instruments are of primary interest in this thesis:

- Generic instruments, which are sometimes grouped into general health profiles and preference-based measures;
- Disease-specific instruments.

Additionally, other types of instruments exist, e.g. those specific to certain symptoms or domains (well-known examples include the Hospital Anxiety and Depression Scale or Beck Depression Inventory).

Generic instruments are intended as general measures of HRQoL that can be applied across different diseases and types of patients, and thus seek to capture a comprehensive range of generally relevant HRQoL domains in their assessment. The most established generic health profile is the Short Form (SF)-36 (Ware, Jr. & Sherbourne, 1992). It consists of 36 questions covering eight domains: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional and mental health. The questions have between two and six response categories. Responses are scaled from 1 (worst) to 100 (best) per domain; additionally, two summary measures called Physical and Mental Component Scores can be derived.

Another example of a general health profile is the Nottingham Health Profile (NHP), which was developed to capture subjective health status as part of health surveys (Coons et al., 2000; Hunt et al., 1980). It contains 38 questions covering six domains: sleep, pain, emotional reactions, social isolation, physical mobility and energy. Each question has two response levels (yes/no); answers are converted to a scale from 1 (best) to 100 (worst) per domain. Floor effects have been observed, i.e. a relatively large proportion of respondents tends to rate themselves as being in full health (which in this case is the lower end of the scale) (Coons et al., 2000).

Preference-based generic measures are of particular interest in a health economic context, as they allow conversion of multidimensional responses to a single index score (utility) that incorporates preferences between health states on a standardised scale ranging from 1 (full health) to 0 (dead). Although information from general health profiles and disease-specific instruments may be converted into preference-weighted index scores by mapping functions for different health states, this raises certain methodological issues which cannot necessarily be resolved in a straightforward manner. The most commonly used and important generic instruments with existing preference scores for health states that will be discussed here are the following:

- EuroQoL EQ-5D;
- SF-6D;
- Health Utilities Index-Mark 3 (HUI-3).

The EuroQoL EQ-5D was developed as a concise multidimensional instrument, covering five attributes: mobility, self-care, usual activities, pain/discomfort and anxiety/depression (Brooks, 1996). In the established 3-level version, each dimension is rated on a scale of 1=no problems, 2=some problems and 3=extreme problems. In an effort to improve sensitivity, a 5-level version has more recently been developed, for which valuation work is ongoing.

Algorithms for converting responses on the EQ-5D to a single index score exist for a range of countries across the world, with the most referenced being the one developed for the UK (Dolan, 1997). Depending on the algorithms used, it is possible to obtain negative utilities (bounded at -1), which denote health states considered worse than death. The EQ-5D is associated with clear ceiling effects, indicating a difficulty of distinguishing between health states close to perfect health (Brazier et al., 2004).

A review of utilities obtained from the EQ-5D in cardiovascular disorders (Dyer et al., 2010) found the instrument to have overall satisfactory psychometric properties in these types of indications, including convergent and discriminative validity, reliability and responsiveness. By nature, it is less sensitive to clinical changes than disease-specific instruments. The dimensions most affected by cardiovascular diseases were pain/discomfort, followed by usual activities and mobility (Dyer et al., 2010).

The SF-6D was developed to allow derivation of utility weights from the SF-36 (Brazier, Roberts & Deverill, 2002). It uses six dimensions of the SF-36 by excluding general health and combining physical and emotional roles in one dimension. In contrast to the EQ-5D, its index scores cannot be negative (Richardson et al., 2014). Moreover, it is associated with floor effects, indicating a lack of sensitivity in distinguishing between the most severe health states (Brazier et al., 2004).

The HUI system uses preferences according to von Neumann-Morgenstern utility theory (discussed in Section 2.2.1.3). The HUI-3 is recommended as the primary version to be used (Drummond et al., Oxford University Press, 2005). It covers eight dimensions of health,

including vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain (Furlong et al., 2001). Similar to the EQ-5D, it is associated with ceiling effects (Hatoum, Brazier & Akhras, 2004) and it is possible to obtain negative utilities (Furlong et al., 2001).

Table 2 summarises key characteristics of the three generic health status classification systems with existing preference scores discussed above. The valuation techniques that have been used for the different instruments are explained further in Section 2.2.1.4. There are clearly marked differences in methods across the instruments, including coverage of different health states, use of different elicitation methods and sources for preferences, as well as different statistical modelling techniques. Therefore, it is natural that these methods result in different utilities even when used to measure the same health states, as has been shown in various studies (e.g. Richardson et al., 2014).

**Table 2: Description of key generic preference-based HRQoL instruments**

	<b>EQ-5D</b>	<b>SF-6D</b>	<b>HUI-3</b>
<b>Main reference</b>	Brooks, 1996	Brazier et al., 2002	Furlong et al., 2001
<b>Number of items (questions)</b>	5	11	8
<b>Dimensions/ attributes</b>	5 - mobility, self-care, usual activities, pain/discomfort, anxiety/depression	6 - physical functioning, role limitations, social functioning, pain, mental health, vitality	8 - vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain
<b>Response levels per question</b>	3 categories (EQ-5D-3L) 5 categories (EQ-5D-5L)	4-6 categories	5 or 6 categories
<b>Number of possible health states</b>	243 (EQ-5D-3L)	18 000	972 000
<b>Valuation techniques used</b>	Time trade-off, visual analogue scale	Standard gamble	Standard gamble, visual analogue scale
<b>Conversion method to single index score</b>	Econometric modelling	Econometric modelling	Multiplicative multi-attribute utility function
<b>Other characteristics</b>	Ceiling effects Negative index scores possible	Floor effects Index score cannot be negative	Ceiling effects Negative index scores possible

Compared to generic instruments, disease-specific measures are inherently more responsive to changes in the studied condition as they are designed to cover areas affected by the disease. However, they usually do not provide comprehensive measures of HRQoL as is required for economic evaluations, nor can the answers readily be converted to preference-weighted index scores.

In AF, there has until recently been a lack of comprehensive disease-specific HRQoL instruments. The Symptom Checklist is a symptom scale focusing on frequency and severity of symptoms and has often been used in AF (Coyne et al., 2005). Recently, there has been development work for at least two new, more comprehensive HRQoL instruments specifically addressing issues in AF: the Atrial Fibrillation-Quality of Life (AF-QoL) questionnaire with 18 questions covering three domains: psychological, physical and sexual activity (Arribas et al., 2010); and the Atrial Fibrillation QualiTy-of-life (AFEQT)

questionnaire with 20 items covering four domains: symptoms, daily activities, treatment concerns and treatment satisfaction (Spertus et al., 2011).

In HF, a number of disease-specific instruments have been developed (Berry & McMurray, 1999; Garin et al., 2009). The most frequently used instrument in clinical trials is the Minnesota Living With Heart Failure Questionnaire (MLWHFQ), which consists of 21 items covering symptoms and signs, physical activity, social interaction, sexual activity, work and emotions. The Chronic Heart Failure Questionnaire (CHFQ) contains 20 items covering four domains: dyspnoea, fatigue, emotion and environmental mastery. More recently, the Kansas City Cardiomyopathy Questionnaire (KCCQ) has emerged as another useful alternative; it comprises 23 items covering physical function, symptoms, social function, self-efficacy and quality of life (Eurich et al., 2006).

### *2.2.1.3 Utilities and the link to QALYs*

The topic of preference-based generic measures leads to other central concepts for this thesis, namely utilities and quality-adjusted life years (QALYs). Utility theory as it is used in health economics is based on von Neumann-Morgenstern's theory of rational decision-making under uncertainty (Drummond et al., Oxford University Press, 2005). This theory is centred on three axioms (normative assumptions) regarding preferences: transitivity, independence and continuity. Transitivity means that if an uncertain prospect  $a$  is preferred over another prospect  $b$ , which in turn is preferred over a third prospect  $c$ , then  $a$  is preferred over  $c$ . Similarly, if there is indifference between  $a$  and  $b$  as well as  $b$  and  $c$ , then the same holds for  $a$  and  $c$ . Independence means that there should be indifference between an uncertain prospect consisting of a single outcome and an equivalent prospect consisting of two outcomes with the same overall probability. Continuity means that if between three outcomes  $x$ ,  $y$  and  $z$ ,  $x$  is preferred over  $y$ , which in turn is preferred over  $z$ , then there exists a probability  $p$  at which there is indifference between experiencing  $y$  with certainty or experiencing the uncertain prospect of either outcome  $x$  with probability  $p$  or outcome  $z$  with probability  $1-p$ .

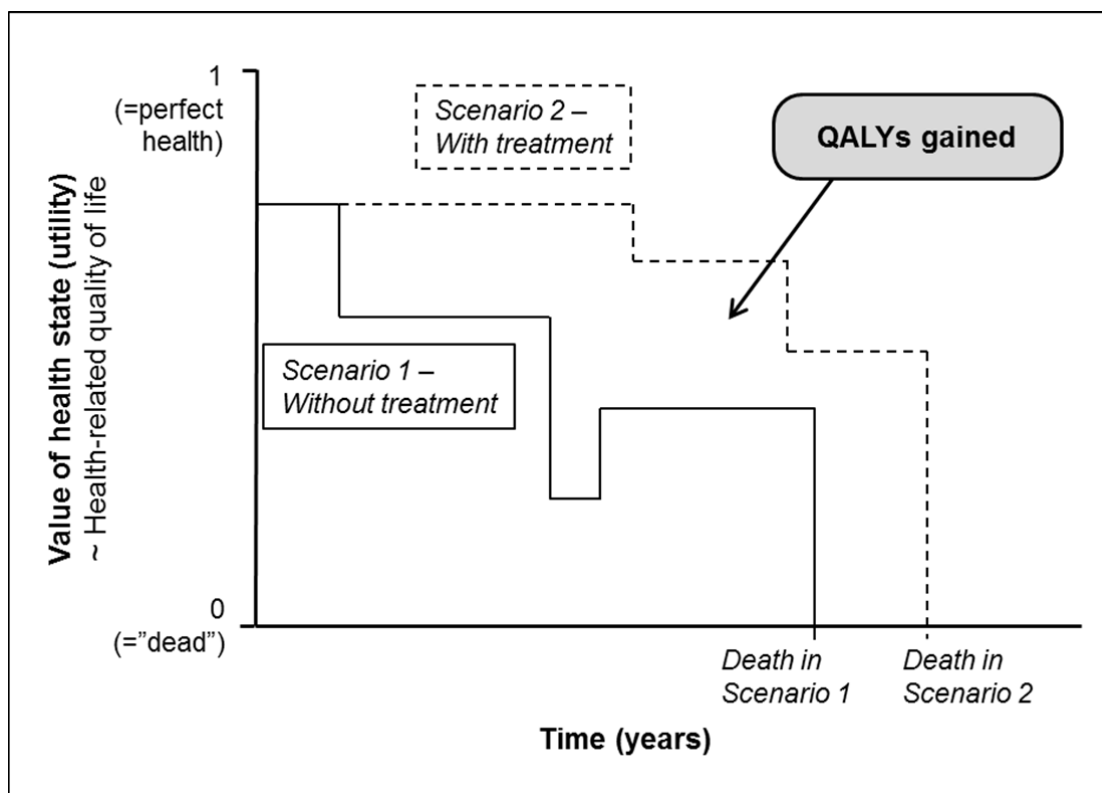
As defined by Drummond and colleagues (Drummond et al., Oxford University Press, 2005), preference is a broader term that encompasses both utilities (obtained for uncertain outcomes) and values (obtained for certain outcomes). There also is a distinction between preferences derived from a scaling exercise (e.g. rating scale, visual analogue scale) or making a choice (e.g. time trade-off, standard gamble). Only the standard gamble (discussed further in Section 2.2.1.4) captures risk attitude and thus is the only method that formally produces utilities as defined by von Neumann-Morgenstern.

Utilities are an essential component for the calculation of QALYs, which are used as a common outcome measure in health economic evaluations (discussed further in Section 2.2.2). The underlying concept of QALYs is that an individual's health states change over time, and that health states can be valued by multiplying their duration with a corresponding quality weight (utility). As opposed to HRQoL, utilities thus assign values to health states.

The overall experience of health can then be quantified by summing up the values for the different health states over time (Weinstein, Torrance & McGuire, 2009).

The QALY weights (utilities) reflect preferences for the relevant health states. They are anchored at 1 (perfect health) and 0 (death or health state considered equal to death in the view of the respondent), meaning that negative QALY weights are theoretically possible. For their use in economic evaluations, QALY weights need to fulfil the properties of an interval scale (Drummond et al., Oxford University Press, 2005). Figure 3 illustrates the concept of QALYs in the context of comparing different treatment alternatives, as well as the role of utilities for obtaining QALYs.

**Figure 3: Concept of quality-adjusted life years and gains from treatment interventions**



Notes: QALY=quality-adjusted life year

From a methodological point of view, QALYs need to meet three conditions to qualify as a von Neumann-Morgenstern utility function (Drummond et al., Oxford University Press, 2005):

- constant proportional trade-off: the proportion of remaining life years that a person would be willing to trade against a defined improvement in quality of life is independent of the amount of remaining life years;
- utility independence: preferences for gambles on quality of life are independent of the the amount of life years and vice versa;

- risk neutrality with respect to time: at a fixed level of quality of life, the utility of additional life years is linear with time.

Relating to these theoretical grounds, there is considerable discussion about the shortcomings of QALYs and possibilities of addressing these to enhance the benefits of this universal outcome measure when comparing health care programmes across different diseases. A lot of the debate can be attributed to the gap between normative utility theory and how people behave and make decisions in reality. In this vein, it has been suggested that researchers may want to investigate “irrationality-adjusted” QALYs (Garrison, Jr., 2009).

Some of the main criticisms of the QALY relate to the strong underlying assumption that health state values are independent of the duration of the health state as well as preceding or following health states (Tsuchiya & Dolan, 2005; Weinstein et al., 2009). There is evidence that suggests that people are not risk neutral with regards to time, in that the value of a health state is related to its duration. This can be observed both in the sense of “maximal endurable time” (Tsuchiya & Dolan, 2005) for severe vs. less severe health states, and in the contrasting phenomenon of coping mechanisms, which involve changes in valuations of the same health state over time. In HRQoL research, the coping phenomenon can be seen to fall under the broader theory of response shift (Sprangers & Schwartz, 1999).

A reflection of the conflict between normative theory and actual behaviour is the discussion around the interpretation of QALYs as belonging to the realm of extra-welfarism, which according to Brouwer and colleagues (Brouwer et al., 2008) is differentiated from welfarism (where social welfare is defined as a function of only individual utility) in the following four ways:

- It allows other outcomes than utilities to be taken into account;
- It allows valuations from others than the affected individuals;
- It allows weightings that are not preference-based;
- It allows comparison of well-being between individuals using a range of dimensions.

Some of these points will be touched upon in the following two subsections.

As part of the discussion around QALYs, several alternatives have been proposed. The DALY (disability-adjusted life year) is the most prominent one due to its use by the WHO as part of its analyses of global disease burden. Compared to QALYs, DALYs focus on health rather than quality of life or well-being by measuring time loss due to premature death and time in less than optimal health (World Health Organization, 2013). DALYs have also been subject to considerable debate, resulting in a revised and simplified methodology in conjunction with the 2010 Global Burden of Disease study. According to the latest methodology, DALYs no longer include age weighting and time discounting, healthy life years lost due to disability are based on prevalence instead of incidence estimates, and they are adjusted for independent comorbidities. In the 2010 Global Burden of Disease study, a reference life expectancy at birth of 86 years has been used. Disability weights

(corresponding to utilities in the QALY context) have been derived as part of a large international study using discrete choice comparisons in individuals from the general population (Salomon et al., 2012). Compared to earlier studies, the new method thus leads to higher number of DALYs lost, with a relatively higher increase amongst younger and older ages. In summary, DALYs can be conceptualised as a measure of population health as opposed to broader aspects of social welfare.

In the context of economic evaluations, the QALY remains at the core of the methodological toolkit, with ongoing discussions and research seeking to identify new methods to address the issues raised by different research groups. There is some consensus around establishing a reference case to move the field forward and to foster wider use outside of resource allocation decisions (Drummond et al., 2009). Interesting research is ongoing e.g. around a revaluation of the UK EQ-5D tariff, which investigates multiple issues around health state valuation, such as duration and timing of health states (including those based on the EQ-5D-5L), methods for elicitation (including discrete choice experiments and methods to evaluate health states worse than death) and modes of administration (Mulhern et al., 2014). Moreover, the EQ-5D-5L development has led to establishment of a standardised international protocol proposing valuation methods to obtain value sets for the expanded number of health states (Devlin & Krabbe, 2013).

#### *2.2.1.4 Methods for obtaining utilities*

There are two different types of methods for obtaining utilities (Drummond et al., Oxford University Press, 2005):

- Directly through preference elicitation techniques, including the standard gamble, time trade-off or scaling techniques (see Green, Brazier & Deverill, 2000; Torrance, 1987);
- Indirectly via generic preference-based HRQoL instruments, such as the EQ-5D, SF-6D and HUI-3 (described in Section 2.2.1.2).

The standard gamble method offers a choice between two alternatives and is directly based on von Neumann-Morgenstern's utility theory. For chronic states, the alternatives consist of either treatment A with two possible outcomes (living in perfect health for  $t$  years with probability  $p$  or immediate death with probability  $1-p$ ), or treatment B with certain outcome of chronic disease state  $i$  for the remaining life ( $t$  years). The probability  $p$  is varied until there is no difference in preference between the two alternatives.

The time trade-off method also involves a choice between two alternatives, but without using probabilities, thus seeking to address potential problems amongst respondents when dealing with the concept of probabilities. For chronic states, the alternatives would be living in health state  $i$  for time  $t$  (corresponding to life expectancy with the condition) until death, or living in full health for time  $x < t$  until death. The time  $x$  is varied until indifference between the two options is achieved.



Lastly, scaling techniques cover e.g. rating scales, category scales or visual analogue scales (VAS). They are generally seen as theoretically inferior to choice-based methods such as standard gamble or time trade-off, because of the inherent risk of scaling bias. However, their advantage lies in being easy and fast to administer; VAS is e.g. also part of the EuroQoL EQ-5D system.

#### *2.2.1.5 Sources for obtaining utilities*

Who should be the source of preferences (or utilities) is a much-debated topic in health economics (Drummond et al., 2009). Traditionally, the general public has been the predominant source for these weights in many countries. One argument supporting this practice has been that in publicly financed health care systems, population-wide funding decisions should also be informed by the preferences of the overall population, which may (or may not) at some point experience the relevant health state(s). Alternatives include asking individuals with experience (past or current) of the relevant health state(s), which often is argued to be more relevant by those involved in clinical decisions. However, this approach may be related to problems of unwillingness to trade length of life against improvements in quality of life amongst people affected by illness or disability, resulting in potential overestimation of utilities.

There is sizable evidence on the effect of the perspective and relevant experience on health state valuations. Based on an analysis of the US valuation for the EQ-5D, there was very limited difference in stated preferences for hypothetical health states between people with or without chronic illnesses (Pickard, Tawk & Shaw, 2013). This was also seen in a meta-analysis of patient and population preferences (Dolders et al., 2006). On the other hand, the meta-analysis also found higher values when comparing preferences derived from patients with current experience of a health state to those from the general population, which would indicate a coping mechanism (or response shift). Therefore, the context of the decision as well as the theoretical or normative standpoint are important when choosing whom to ask (see e.g. Burstrom et al., 2014). For example, in treatment decisions, the patient could be seen as the most informed about the relevant health states, while for preventative programmes, healthy individuals part of the general public may be more relevant (although they also may have difficulties envisaging hypothetical health states). Moreover, different health technology assessment (HTA) bodies take different stances on this, with e.g. the National Institute for Health and Care Excellence (NICE) in the UK preferring values obtained from the general population (National Institute for Health and Care Excellence, 2013), while the Swedish Dental and Pharmaceutical Benefits Agency (TLV) favours experience-based values (Tandvårds- och Läkemedelsförmånsverket, 2003).

### **2.2.2 Economic evaluations**

In this section, the main principles for economic evaluations and cost-effectiveness analysis are discussed. In summary, economic evaluations can be defined as “the comparative analysis of alternative courses of action in terms of both costs and their consequences” (Drummond et

al., Oxford University Press, 2005). As part of efforts to make systematic and rational judgements on how best to allocate limited resources between a growing number of therapeutic alternatives, economic evaluations have become an important element of the decision-making process by many HTA authorities across the world.

#### *2.2.2.1 Types of economic evaluations*

There are four main types of economic evaluations, which differ in how they define and measure the consequences, or effects, of the studied treatment alternatives:

- Cost-minimisation analysis: the consequences are considered to be the same, and only costs are compared. This type of evaluation can be considered as a more limited analysis of only costs, given that the alternatives do in fact have the same effect;
- Cost-effectiveness analysis: the consequences are made up of only a single effect that is the same for both alternatives. The effect is measured in natural units, e.g. units of lipid reduction (mmol/L), life years gained, number of events avoided, etc.;
- Cost-utility analysis: the consequences can be made up of either single or multiple effects, which do not need to be the same for both alternatives. The outcomes are usually measured in QALYs (see Section 2.2.1.3);
- Cost-benefit analysis: this widens the cost-utility concept to the measurement of consequences in monetary units.

The focus in the literature, as well as this thesis, is generally on cost-effectiveness and cost-utility analysis. In particular, cost-utility analysis using QALYs is often recommended as the “reference case” by different authorities or research groups (e.g. National Institute for Health and Care Excellence, 2013; Siegel et al., 1997).

#### *2.2.2.2 The question of perspective*

Regardless of the type of analysis and specific method chosen, all economic evaluations should start with a formulation of the decision problem – namely what is being compared and from which perspective. The most appropriate comparison from the viewpoint of a decision-maker is between a new treatment (which may be a pharmaceutical, medical device, vaccination programme or other health care intervention) and standard practice (which often is something else than “no treatment”).

The analysis perspective is essential, as it determines which costs and effects are included in the evaluation. The two main approaches are a societal perspective or a payer/health care perspective. In a societal perspective, all effects (in terms of costs and health outcomes) of an intervention should be included, regardless of where they occur. In a payer/health care perspective, costs that are not paid by the health insurance system are not included in the calculations. The stance on which perspective to use differs across bodies, with e.g. the US Panel on Cost-effectiveness in Health and Medicine designating the societal perspective as the “reference case” (Siegel et al., 1997), while e.g. NICE in the UK considers the payer and personal social services perspective to be most relevant (National Institute for Health and

Care Excellence, 2013), or the Canadian Agency for Drugs and Technologies in Health (CADTH) that of the publicly funded health care system (Canadian Agency for Drugs and Technologies in Health, 2006).

The types of costs can be grouped into the following categories:

- Direct health care (medical) costs, e.g. hospital care, outpatient visits, diagnostic tests, drug costs;
- Direct non-medical costs, e.g. transportation costs, social services;
- Indirect costs, which cover productivity losses, e.g. short-term work absence, long-term sick leave, early retirement;
- Informal care (cost of caregiver time) – depending on school of thought, these can be measured either as a direct (replacement) or indirect (productivity) costs.

To obtain costs, resources should be valued at their market prices, which are deemed a suitable representation of opportunity cost, i.e. the value of the benefits that cannot be realised because the resource is not available for its best alternative use (Drummond et al., Oxford University Press, 2005).

Another important component of economic evaluations is the time horizon during which the costs and effects are measured. Ideally, it should be sufficiently long to capture all relevant costs and effects, also those resulting from the given condition in the future.

#### *2.2.2.3 Incremental cost-effectiveness ratios*

In all types of economic evaluations (except cost-minimisation analysis), the concept of an incremental cost-effectiveness ratio (ICER) plays a central role. The ICER is defined as the ratio between the incremental costs and the incremental effects of an intervention A vs. its comparator B:

$$\text{ICER} = (\text{Costs}_A - \text{Costs}_B) / (\text{Effects}_A - \text{Effects}_B).$$

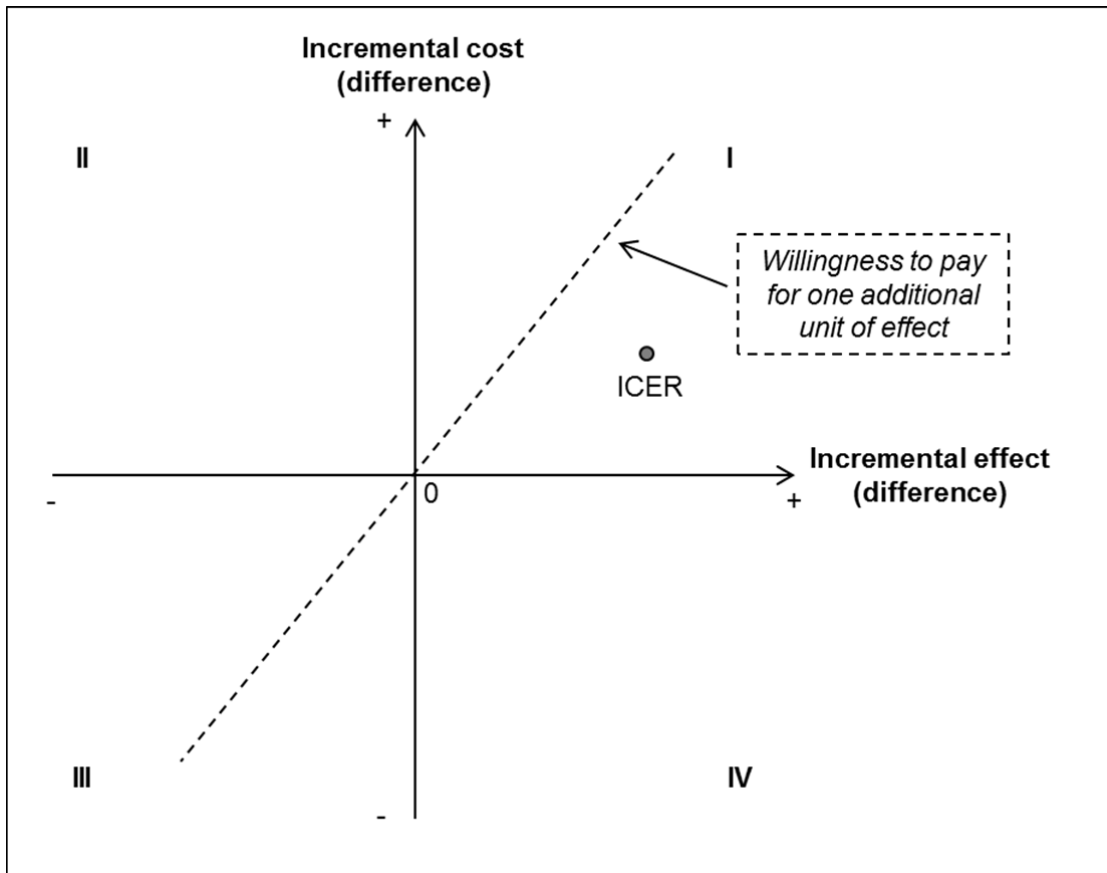
The ICER can be illustrated in relation to the possible results that may be obtained when comparing two mutually exclusive treatment alternatives. The differences in costs and effects can be plotted on the so-called cost-effectiveness plane (Figure 4).

Depending on which quadrant the results lie in, different decisions may be taken:

- Quadrant I: A costs more than B and is more effective. The decision whether it can be considered cost-effective or not depends on the willingness to pay (WTP) per unit of effect gained – if the ICER is below a defined threshold WTP (represented by a dotted line in the graph), then A would be cost-effective;
- Quadrant II: A costs more than B and is less effective, i.e. should not be adopted. In this case, A is said to be dominated by B (or B is the dominant alternative);
- Quadrant III: A costs less and is less effective than B. As in Quadrant I, the decision depends on whether the ICER falls below a defined WTP threshold;

- Quadrant IV: A costs less than B and is more effective, i.e. should be adopted. In this case, A is the dominant alternative.

**Figure 4: The cost-effectiveness plane**



*Notes: ICER=incremental cost-effectiveness ratio*

#### 2.2.2.4 Trial-based analyses vs. decision-analytic modelling

Obtaining reliable ICER estimates is at the heart of numerous health economic guidelines, including those by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (Caro et al., 2012; Ramsey et al., 2005). Attention needs to be given to conceptualising the decision problem, development of analytical strategy, derivation or estimation of input values and assessment of uncertainty. It should be noted that various analytical approaches and data tend to be used as complementary pieces of evidence considered by decision-makers. There is no single “one size fits all” solution, which reflects the complexities of reality in this field. There are e.g. potentially many different sources for costs and effects, both with and without the new treatment. Also, the chosen perspective and time horizon tend to require combinations of different sources and assumptions.

There has traditionally been a differentiation between cost-effectiveness (or cost-utility) analyses performed alongside clinical trials, so-called trial-based analyses, and analyses based on decision-analytic models. Trial-based analyses are usually performed on the basis of patient-level data directly collected as part of a (mostly randomised controlled) clinical trial,

focusing specifically on the treatment options studied and the costs and effects observed during the trial. The availability of patient-level data represents a valuable advantage for statistical analyses, as it allows addressing skewed cost data and obtaining confidence intervals (uncertainty estimates) for ICERs. When based on a randomised controlled trial, these analyses also have the benefit of having high internal validity, as potential confounders influencing treatment decisions have been removed. Nevertheless, there are a number of potential caveats to take into consideration (as outlined in e.g. Drummond et al., Oxford University Press, 2005 and Ramsey et al., 2005):

- Efficacy vs. effectiveness: clinical trials are designed to measure efficacy, i.e. the therapeutic effect that can be attributed purely to treatment, thus requiring control of many external factors. This means that data need to be complemented by other evidence on real-world effectiveness;
- Selection of patient population: inclusion criteria for clinical trials tend to be rather specific, potentially leading to difficulties when seeking to generalise trial results to other patient groups;
- Relevance of comparator therapy: many randomised controlled trials use placebo as the comparator, while more often than not, standard of care in clinical practice includes active treatment(s);
- Protocol-driven costs and outcomes: clinical trials have predefined schedules of assessment, which may lead to additional costs; better compliance than in clinical practice may also lead to different outcomes. Related issues are to what degree the primary outcome (e.g. a combined endpoint) is informative for a health economic evaluation, and whether the duration of follow-up is sufficiently long to capture events of interest after the endpoint has been achieved;
- Multinational clinical trials: differences in patient populations, treatment practices and health care systems need to be tested for and potentially taken into account (e.g. through hierarchical modelling, see Manca, Sculpher & Goeree, 2010);
- Censored cost data: contrary to time to event data, cost data may not be censored at random in clinical trials, as patients accrue costs at different rates depending on their disease severity. Different methods exist to take this into account, e.g. assuming random censoring (Ramsey et al., 2005) or using non-parametric methods (Raikou & McGuire, 2004).

“Pure” trial-based analyses may be extended to incorporate some modelling approaches, in particular regarding extensions beyond the trial duration to measure long-term outcomes of interest in economic evaluations. Prediction of survival beyond the trial duration may be done through parametric, semi-parametric or non-parametric methods (Gerdtham & Zethraeus, 2003); in all cases, the impact of assumptions should be tested through sensitivity analyses.

Decision-analytic modelling seeks to address the issues encountered when using single clinical trials as the basis for health economic evaluations through e.g. combination of data from different sources (such as meta-analysis for treatment effects) and extrapolation of costs

and effects beyond the trial setting. Models use a mathematical framework to approximate a representation of reality, which requires simplification and assumptions. Many different modelling techniques exist, depending on the complexity of the decision problem and required model structure (Caro et al., 2012):

- Cohort models: involve transitions between health states by modelling cohorts of patients; these include simpler decision trees and more complex Markov models;
- Microsimulation models: involve transitions between health states by modelling individuals; these allow more complex patient histories to be taken into account than practicable with cohort models;
- Discrete event simulation: allows time to event to be modelled stochastically, i.e. not using fixed time intervals;
- Dynamic disease models: suitable for capturing transmission mechanisms in e.g. infectious diseases;
- Hybrid models: combinations of different techniques may be used, e.g. combining health states and events to allow flexible handling of time dimension.

Given the potential complexity of models, clarity in design, implementation and communication is highly important to produce credible and valuable evidence for decision-makers. Best practice modelling guidelines provide detailed recommendations around these topics, including model calibration, validation of results with clinical and health economic experts as well as existing evidence, and transparency in reporting (Caro et al., 2012).

#### *2.2.2.5 Evaluating uncertainty in economic evaluations*

Uncertainty around cost-effectiveness results can stem from a variety of sources, including methodological choices, data inputs, a need to extrapolate results over longer time horizons or to apply the results in another setting. Sensitivity analyses can illustrate the impact of different assumptions regarding structure and input parameters on results. Uncertainty in input parameters can also be assessed using probability distributions for key inputs, including costs, outcomes and treatment effects (so-called probabilistic sensitivity analysis). Where patient-level data is available, statistical methods such as bootstrapping can be used to capture sampling variation.

The results of probabilistic sensitivity analyses can be presented in a number of ways, including confidence intervals around the ICER and cost-effectiveness acceptability curves. In a Bayesian framework, a cost-effectiveness-acceptability curve shows the probability that a treatment is cost-effective compared to its comparator at different thresholds of WTP for additional units of effect (Drummond et al., Oxford University Press, 2005; Zethraeus et al., 2003). It is thus a useful illustration of the uncertainty underlying a decision. Although its concept is often illustrated by a smooth increasing curve asymptotically approaching 1, it is in fact not a cumulative distribution function and can take many functional forms (Fenwick, O'Brien & Briggs, 2004). Related to this concept is a wealth of studies seeking to establish the WTP for a QALY (also termed cost-effectiveness thresholds) as the most common

outcome measure in economic evaluations (see e.g. review by Ryen & Svensson, 2014), and a debate about whether and how these thresholds should be applied (e.g. Eichler et al., 2004; Gafni & Birch, 2006).

### **2.2.3 Risk stratification for prognosis and management**

In this section, the focus lies on the use of HRQoL in a clinical decision-making context. As outlined earlier, HRQoL is an important component of disease burden, and it has been described and analysed with different research objectives in mind. Apart from being an outcome measure for a disease (and potentially economic evaluations), it can also be studied as a potential driver of other outcomes, such as disease progression or death. These kinds of prediction models aim to classify patients according to different levels of risk for an outcome of interest, and serve to target the intensity of treatment and follow-up. They thus present an important tool for clinical management, and can ultimately aid in reducing disease burden. Risk prediction can also be useful for identifying patient groups most likely to benefit from new treatments, e.g. as part of clinical trials. These models can cover a range of different factors, often focusing on demographic, clinical and laboratory values. For uptake in clinical practice, they need to be simple to use, be based on standard available data, have been validated in external data sets and have shown good predictive ability (calibration).

The interest in identifying a simple prognostic measure in several chronic diseases is reflected in the sizeable number of studies that have investigated the impact of self-reported health status (using a single question or more elaborate HRQoL instruments) on e.g. mortality. For example, in CHF and coronary artery disease, physical health status independently impacted mortality and hospitalisation outcomes in the majority of 34 studies covered by a systematic review (Mommersteeg et al., 2009). Other examples include oncology, where HRQoL has been shown to be a predictor of mortality in numerous indications (Steel et al., 2014); in arthritis, physical and mental health components have been shown to be significant drivers of resource utilisation and mortality (Singh et al., 2005).

In AF and CHF, risk prediction models commonly used today do not include HRQoL, but focus mostly on clinical measures. A key prognostic model in AF is the CHA<sub>2</sub>DS<sub>2</sub>-VASc score which stratifies patients according to their probability of experiencing a stroke based on several risk factors. As illustrated in Figure 5, based on the resulting risk score, different levels of antithrombotic treatment are recommended (e.g. ESC guidelines).

**Figure 5: Risk stratification of AF patients for antithrombotic treatment using the CHA<sub>2</sub>DS<sub>2</sub>-VASc score (adapted based on ESC 2010 guidelines - Camm et al., 2010)**

Risk factors	Score	$\Sigma$	CHA <sub>2</sub> DS <sub>2</sub> -VASc score and associated stroke risk	Recommended antithrombotic treatment
Congestive heart failure/left ventricular dysfunction	1		0 (no risk factors) Annual stroke rate: 0%	Aspirin or no antithrombotic therapy
Hypertension	1		1 (one clinically relevant non-major risk factor) Annual stroke rate: 1.3%	Oral anticoagulant or aspirin
<b>Age<math>\geq</math>75 years</b>	<b>2</b>		$\geq$ 2 (one <b>major risk factor</b> or $\geq$ 2 clinically relevant non-major risk factors) Annual stroke rate: 2.2-15.2%	Oral anticoagulant
Diabetes	1			
<b>Stroke/TIA/thromboembolism</b>	<b>2</b>			
Vascular disease	1			
Age 65-74 years	1			
Sex category (i.e. female)	1			

Notes: TIA=transient ischaemic attack

In CHF, several prognostic models for predicting mortality risk have been developed. A well-known example is the Seattle Heart Failure model, which predicts mortality over 1-3 years (Levy et al., 2006). It was originally derived based on 1125 patients with HFREF (LVEF<30%) and subsequently validated in a broad range of populations; however, a recent systematic review found the external validations to indicate poor to acceptable discrimination and questionable calibration (Alba et al., 2013). Moreover, it covers 24 clinical, pharmacological, device and laboratory variables, which may not always be available in routine clinical care, e.g. lymphocyte counts or uric acid.

More recently, a new risk score (MAGGIC) has been developed for 3-year CHF mortality based on a meta-analysis of data from 39 372 patients in 30 studies, covering a comprehensive spectrum of patient types, including those with HFPEF (Pocock et al., 2013). The risk score is based on 13 prognostic markers that are deemed to be readily available in clinical practice: age, LVEF, NYHA class, serum creatinine, diabetes, beta-blockers, systolic blood pressure, body mass index, time since diagnosis, smoking status, chronic obstructive pulmonary disease, gender and use of angiotensin converting enzyme inhibitors/angiotensin receptor blockers. External validation has to date been performed in the Swedish Heart Failure Registry covering 51 043 patients, indicating modest predictive accuracy for lower and higher risk patients, but overall confirming the risk score's discriminatory power (Sartipy et al., 2014).

Although not yet externally validated, the HF-ACTION prediction model includes HRQoL as a prognostic factor for hospitalisation or death (O'Connor et al., 2012). The model was derived in a sample of 2331 ambulatory patients with LVEF $\leq$ 35% and NYHA classes II-IV who received current evidence-based treatment (of note, this may not be the case in some of the studies part of MAGGIC, see Sartipy et al., 2014). KCCQ symptom stability, indicating no symptoms or no change in symptoms over the last two weeks, was the second most important predictor in the model. Interestingly, change in either direction increased the risk of a negative outcome compared to symptom stability, with a hazard ratio of 1.91 (95% CI:



1.62–2.27) for worsened symptoms and a hazard ratio of 1.26 (95% CI: 1.11–1.43) for improved symptoms (O'Connor et al., 2012).

## **2.3 REVIEW OF THE LITERATURE**

A focused literature review was performed for each of the relevant subtopics covered by this thesis. The selection was based on papers with a primary focus on the research topic of interest. The search was performed in PubMed and considered publications from the year 2000 onwards. The following sections present selected findings from studies deemed interesting and relevant.

### **2.3.1 Health-related quality of life and utilities in atrial fibrillation**

Many of the studies on factors influencing HRQoL in AF have been part of randomised controlled trials that have investigated different treatment strategies, in particular rate vs. rhythm control. While a range of different HRQoL instruments have been used in AF overall, most of these studies have used generic HRQoL instruments, in particular the SF-36 or abbreviated versions.

Table 6 summarises key information from studies on drivers of HRQoL in AF. The studies have been selected to provide information from samples with more than 100 patients, and because they investigated a range of possible HRQoL drivers. The two largest studies have been performed in North America and the UK, respectively. Reynolds and colleagues (Reynolds et al., 2006) reported from a prospective observational registry of 963 patients with new onset AF, using the SF-12. Roalfe and colleagues (Roalfe et al., 2012) used both the EQ-5D and SF-36 in an aged AF population of 1762 patients who were followed for stroke prevention in primary care.

Variables that previously have been identified as independent determinants of HRQoL include gender, age, disease duration, AF symptoms, LVEF, NYHA class, valvular heart disease, depression, anxiety, degree of medication use (including beta-blockers, antiarrhythmic drugs), disability, comorbidities (including diabetes, hypertension, chronic obstructive pulmonary disease) and, where information over time was available, adverse events during follow-up. Type of AF (paroxysmal, persistent or permanent) has only been shown to be significant in some studies, which may partly be due to the specific populations and research questions. In general, randomised controlled trials comparing rhythm vs. rate control (e.g. AFFIRM, PIAF, RACE) have not shown any significant difference in overall HRQoL improvement with treatment between the types of therapy measured using the generic SF-36 instrument, although some differences between individual domains have been observed. It has been noted that this could be either due to lack of treatment effect in terms of health status, or insufficient sensitivity of the SF-36 to capture any smaller changes seen with pharmacological treatment (Coyne et al., 2005; Spertus et al., 2011).

Apart from the utility functions in Paper I, health state utilities have only been published for selected disease states in AF. Based on an observational registry of new onset AF (Reynolds,

Morais & Zimetbaum, 2010), utilities derived from the SF-6D were compared between patients with and without hospitalisations during the first year. Patients not hospitalised had a mean utility of 0.81, compared to 0.77 for patients hospitalised for any cause ( $p<0.01$ ). This relationship was confirmed when using a mixed linear regression model for the full 2.5-year follow-up period. Recently, a cross-sectional multinational survey of about 10 000 patients with AF (Steg et al., 2012) reported utilities based on the EQ-5D for AF control (defined as being in sinus rhythm or having AF with a resting ventricular rate  $\leq 80$  beats per minute). Controlled patients had a median utility score of 0.78 compared to 0.73 for uncontrolled patients ( $p<0.001$ ).

**Table 6: Selected studies on HRQoL drivers in AF**

Publication	Country	Setting	Study design	Sample size	Patient characteristics at baseline	HRQoL measure	Main results
<b>Dorian et al., 2002</b>	Canada	Clinical centres	Prospective sub-study of RCT evaluating AADs (CTAF)	294	Paroxysmal or persistent AF Mean age 65 years 89% had LVEF>50% NYHA class I-II	SF-36	Correlates of SF-36: age, gender, NYHA class
<b>Hagens et al., 2004a</b>	The Netherlands	Clinical centres	Sub-study of RCT on rate control vs. cardioversion (RACE)	352	Persistent AF Mean age 68 years NYHA class I-III	SF-36	Baseline correlates: gender, age, AF duration, AF symptoms, NYHA class, fractional shortening Determinants of change in HRQoL at study end (24 or 36 months): age, AF symptoms, AF duration, sinus rhythm at end of follow-up
<b>Reynolds et al., 2006</b>	US, Canada	Clinical centres	Observational registry	963	New onset AF Mean age 66 years	SF-12	Physical and/or mental summary scores over 12 months driven by: NYHA class, gender, age, valve disease, prior MI, hypertension, initial AAD, initial warfarin, COPD, asthma, diabetes, hemiplegia, CTD
<b>Peinado et al., 2010</b>	Spain	Routine clinical practice	Prospective observational study	341	Paroxysmal, persistent or permanent AF Mean age 60 years 12% LV systolic dysfunction NYHA class I-IV	AF-QoL	Total HRQoL score influenced by: NYHA class, number of emergency room visits during last year AF type only significant in psychological dimension
<b>Groenveld et al., 2011</b>	The Netherlands	Clinical centres	Sub-study of RCT on rate control strategies (RACE II)	437	Permanent AF Mean age 68 years Mean LVEF 53% NYHA class I-III	SF-36	Baseline correlates: gender, AF symptoms, diabetes Determinants of change in HRQoL at study end (2-3 years): age, gender, AF symptoms, LVEF, septum thickness, diabetes, beta-blockers, HF hospitalisation, major bleeding, arrhythmic events during study
<b>Roalfe et al., 2012</b>	UK	Primary care	RCT of stroke prevention (BAFTA)	1762	Mean age 82 years	EQ-5D, SF-12	HRQoL scores independently associated with gender, greater medication use and disability (Rankin scale)
<b>Akintade et al., 2013</b>	US	Hospital clinics	Cross-sectional observational study	150	Paroxysmal, persistent or permanent AF; AFL Mean age 66 years Mean LVEF 48%	SF-36	HRQoL summary scores determined by gender, depression, anxiety, AFL, permanent AF

Notes: AAD=antiarrhythmic drug, AF=atrial fibrillation, AFL=atrial flutter, AF-QoL=Atrial Fibrillation-Quality of Life, BAFTA=Birmingham Atrial Fibrillation Treatment of the Aged, COPD=chronic obstructive pulmonary disease, CTAF=Canadian Trial of Atrial Fibrillation, CTD=connective tissue disease, HF=heart failure, HRQoL=health-related quality of life, LVEF=left ventricular ejection fraction, MI=myocardial infarction, NYHA=New York Heart Association, RACE=RAte Control versus Electrical cardioversion, RACE II=RAte Control Efficacy in permanent atrial fibrillation: A comparison between lenient and strict rate control II, RCT=randomised controlled trial

### **2.3.2 Economic evaluations in atrial fibrillation and the impact of utilities**

Table 7 summarises key information from selected cost-effectiveness evaluations of pharmacological therapies in AF. Studies investigating treatments for stroke prevention have been excluded. With the exception of the study by Marshall and colleagues (Marshall et al., 2004), only studies reporting ICERs have been listed.

AFFIRM and RACE were two major trials investigating rhythm vs. rate control in recurrent and persistent AF, respectively, both of which showed non-inferiority of rate control in terms of survival. In AFFIRM (Wyse et al., 2002), rate control therapy was associated with a significantly lower risk of hospitalisation and selected adverse events. In RACE (Van Gelder et al., 2002), the primary endpoint comprised cardiovascular morbidity and mortality, with the majority of endpoints occurring numerically more frequently with rhythm control. Reflecting these clinical results, rate control has also been shown to be cost-saving or dominant in subsequent cost-effectiveness analyses.

Ablation has been compared to antiarrhythmic therapy in several studies. It is interesting to note the different methods used for utilities in these evaluations. Data were based on a combination of relatively different sources and assumptions, reflecting the lack of utility data in a comprehensive sample of AF patients at the time. As pointed out by McKenna and colleagues (McKenna et al., 2009), the necessity to use different sources introduced significant uncertainty around this model parameter.

Most recently, a number of cost-effectiveness evaluations have been performed for dronedarone to support reimbursement applications in different markets. They have used different comparators and modelling frameworks, which in particular has informed the choice of data source for event rates (trial-based or from indirect comparisons). In terms of utilities, these studies have used values derived from the Euro Heart Survey (Paper I).

Drivers of cost-effectiveness results are often related to the risk of events with or without treatment and the associated benefits in terms of costs or utility gains. In the study by McKenna and colleagues (McKenna et al., 2009), results were sensitive to assumptions around utilities assigned to health states and in particular the duration of HRQoL benefits following treatment. In the study by Reynolds and colleagues (Reynolds et al., 2009), results were most influenced by relative utility weights of different health states following therapy, cost of ablation and the modelling time horizon. Treatment duration was found to be a key driver in both models comparing dronedarone vs. standard of care (Akerborg et al., 2012; Reynolds et al., 2013). In the models comparing dronedarone against other antiarrhythmic drugs, results were most sensitive to assumptions around mortality benefits of and the incremental survival benefit with dronedarone (McKenna et al., 2012; Nilsson et al., 2013).

**Table 7: Selected cost-effectiveness evaluations of pharmacological interventions in AF (not including stroke prevention)**

Publication	Country (costing)	Intervention	Patient population	Study type	Time horizon and perspective	Outcome measures and utility sources	Main results (base case)
<b><i>Rhythm vs. rate control</i></b>							
<b>Hagens et al., 2004b</b>	The Netherlands	Rate control drugs vs. rhythm control (electrical cardioversion and AADs)	Patients with persistent AF or AFL as in RACE	Prospective sub-study of RACE; cost-minimisation and cost-effectiveness analysis	Mean follow-up 2.3 years Societal	Event avoided (cardiovascular morbidity and mortality)	Rate control: Cost savings €898 per patient Cost savings per event avoided €24944
<b>Marshall et al., 2004</b>	US	AADs vs. rate control drugs	Patients similar to those in AFFIRM (mostly recurrent AF)	Retrospective economic evaluation based on AFFIRM	Mean follow-up 3.5 years Third-party payer	Survival (LYG)	Rate control: Mean survival gain 0.08 years Cost savings \$5077 per patient
<b>Perez et al., 2011</b>	US	Rate vs. rhythm control	Patients with persistent or paroxysmal AF with heart failure	Markov model	Lifetime Third-party payer	QALYs Published index scores for chronic conditions	Rate control dominant
<b><i>Ablation vs. pharmacological treatment</i></b>							
<b>Chan et al., 2006</b>	US	Left atrial catheter ablation vs. rate control drugs or amiodarone	Hypothetical cohort of AF patients at low-moderate risk of stroke	Markov model	Lifetime Societal	QALYs Utilities largely based on assumptions	ICER range \$28700-\$98900/QALY depending on stroke risk and age Amiodarone dominated by rate control
<b>Reynolds et al., 2009</b>	US	Radiofrequency catheter ablation vs. AAD	Drug-refractory paroxysmal AF	Markov model	5 years Societal	QALYs Utilities from FRACTAL registry, prospective cohort and A4 trial	ICER \$51431/QALY
<b>McKenna et al., 2009</b>	UK	Radiofrequency catheter ablation vs. AAD (amiodarone)	Paroxysmal AF	Decision analytic model based on Bayesian evidence synthesis	Lifetime Health service	QALYs Utilities based on decrements from UK general population	ICER range £7763-£7910/QALY depending on stroke risk

<b><i>Dronedarone vs. other drugs</i></b>							
<b>Akerborg et al., 2012</b>	Canada, Italy, Sweden, Switzerland	Dronedarone on top of SOC vs. SOC alone	Paroxysmal or persistent AF or AFL	Microsimulation model, mainly based on patient-level data from ATHENA	Lifetime, treatment 21 months Payer (NHS)	Survival (LYG) QALYs Utilities based on Euro Heart Survey	ICER range €5828-€14970/QALY across countries
<b>McKenna et al., 2012</b>	UK	Dronedarone vs. SOC, amiodarone, sotalol or class 1c agents	Paroxysmal or persistent AF or AFL	Discrete event model, based on data from 4 dronedarone trials and indirect comparisons	Lifetime NHS and PSS	QALYs Utilities based on Euro Heart Survey	ICER range £1692-£18955/QALY across different patient populations and comparators
<b>Reynolds et al., 2013</b>	US	Dronedarone on top of SOC vs. SOC alone	Paroxysmal or persistent AF or AFL	Microsimulation model, mainly based on patient-level data from ATHENA	Lifetime, treatment 21 months Payer	Survival (LYG) QALYs Utilities based on Euro Heart Survey	ICER \$19520/QALY
<b>Nilsson et al., 2013</b>	Canada, Italy, Sweden, Switzerland	Dronedarone on top of SOC vs. amiodarone, sotalol or flecainide on top of SOC	Paroxysmal or persistent AF or AFL	Microsimulation model, mainly based on patient-level data from ATHENA	Lifetime Payer (NHS)	Survival (LYG) QALYs Utilities based on Euro Heart Survey	ICER range €2290-€6140/QALY across countries and comparators

*Notes: A4=Catheter Ablation versus AntiArrhythmic drugs for Atrial fibrillation, AAD=antiarrhythmic drug, AF=atrial fibrillation, AFFIRM=Atrial Fibrillation Follow-up Investigation of Rhythm Management, AFL=atrial flutter, ATHENA=A placebo-controlled, double-blind, parallel arm Trial to assess the efficacy of dronedarone 400 mg bid for the prevention of cardiovascular Hospitalization or death from any cause in PatiENts with Atrial fibrillation/atrial flutter, ICER=incremental cost-effectiveness ratio, LYG=life years gained, NHS=national health service, PSS=personal social services, QALY=quality-adjusted life year, RACE=RAte Control versus Electrical cardioversion, SOC=standard of care*

### **2.3.3 Health-related quality of life and utilities in chronic heart failure**

Determinants of HRQoL in HF have been investigated in a number of studies using both generic and disease-specific instruments, covering patients from hospital and outpatient settings. The largest samples have been based on the CARE-HF trial (Calvert, Freemantle & Cleland, 2005) and the outpatient INCA study (de Rivas et al., 2008), including 813 and 2161 patients, respectively. These two studies, as well as a study in 293 patients with emergency admissions (Holland et al., 2010) and a randomised controlled trial on iron deficiency (Gutzwiller et al., 2013), have used the EQ-5D. Other commonly used HRQoL instruments include the SF-36, MLWHFQ and KCCQ.

Variables that previously have been found to be independent determinants of HRQoL in CHF include: age, gender, NYHA class, LVEF, comorbidities (in particular chronic obstructive pulmonary disease and diabetes), medical history, renal function, body mass index, functional measures, symptoms, medications (including aldosterone antagonists, antiarrhythmic drugs), depression and sociodemographic variables.

Table 8 summarises key information from studies on drivers of HRQoL in CHF. The studies have been selected to provide information from samples with more than 200 patients, and with the goal of covering a broad range of possible HRQoL drivers.

**Table 8: Selected studies on HRQoL drivers in CHF**

Publication	Country	Setting	Study design	Sample size	Patient characteristics at baseline	HRQoL measure	Main results
<b>Calvert et al., 2005</b>	International	Clinical centres	RCT on cardiac resynchronisation therapy (CARE-HF)	813	Mean age 65 years Mean LVEF 26% NYHA III-IV	EQ-5D, MLWHFQ	HRQoL scores significantly related to gender and NYHA class
<b>Lee et al., 2005</b>	Hong Kong	Hospital admissions	Cross-sectional	227	Mean age 77 years NYHA class I-IV	Chronic Heart Failure Questionnaire	HRQoL score significantly correlated with HADS score, health perception, NYHA, educational level
<b>Gott et al., 2006</b>	UK	General practice	Cross-sectional survey	542	Age >60 years NYHA class I-IV	SF-36, KCCQ	HRQoL predictors: gender, NYHA class, depression, socioeconomic group, age, number of comorbidities
<b>Franzen et al., 2007</b>	Sweden	Hospital register	Cross-sectional survey	357	Mean age 79 years Self-rated NYHA I-IV	SF-12, MLWHFQ	Predictors SF-12 physical component summary: NYHA, age, gender, diabetes, respiratory diseases
<b>de Rivas et al., 2008</b>	Spain	Outpatient clinics	Cross-sectional	2161	Mean age 71 years Mean LVEF 48% NYHA class I-IV	EQ-5D, MLWHFQ	Drivers EQ-5D VAS: age, gender, NYHA, LVEF, renal dysfunction, haemoglobin, COPD, diabetes, follow-up in primary care; MLWHFQ: age, gender, NYHA, LVEF, renal dysfunction, serum potassium, COPD, diabetes, hypertension, valve disease
<b>Faller et al., 2009</b>	Germany	Out-patients	Prospective cohort study	206	Mean age 64 years 52% systolic HF NYHA I-IV	KCCQ, Patient Health Questionnaire depression module	Depression influenced psychological and physical KCCQ domains; NYHA class influenced only physical domain
<b>de Leon et al., 2009</b>	US	Out- and inpatient screening	RCT on behavioural intervention	695	Mean age 63 years 24% diastolic HF NYHA class II-III	SF-36 Physical Functioning, QoL Index-Cardiac Ver.	SF-36 scores associated with: age, race, education, number of comorbidities, NYHA class, depression, HF symptoms and functional characteristics
<b>Holland et al., 2010</b>	UK	Emergency hospitalisation	Observational cohort part of RCT (pharmacist advice)	293	Mean age 77 years NYHA class I-IV (self-assigned)	EQ-5D, MLWHFQ	Significant predictors of MLWHFQ at 6 months: self-assigned NYHA class, 3-month MLWHFQ scores, antiarrhythmic drug use, BMI
<b>Peters-Klimm et al., 2010</b>	Germany	Primary care	Cross-sectional	318	Mean age 69 years Mean LVEF 35% NYHA class I-IV	SF-36, KCCQ	Determinants of (some) HRQoL scales: practice list size, age, social class, NYHA class, COPD, history of CABG surgery, multimorbidity, aldosterone antagonist, self-care, depression
<b>Gutzwiller et al., 2013</b>	International	Clinical centres	RCT on iron deficiency	459	Mean age 68 years Mean LVEF 32% NYHA class II-III	EQ-5D, KCCQ	EQ-5D VAS and KCCQ at 24 weeks significantly associated with: treatment group, NYHA class, eGFR, stroke, functional test, country of residence

Notes: BMI=body mass index, CABG=coronary artery bypass grafting, CARE-HF=Cardiac REsynchronisation in Heart Failure, COPD=chronic obstructive pulmonary disease, eGFR=estimated glomerular filtration rate, HADS=Hospital Anxiety and Depression Scale, HF=heart failure, HRQoL=health-related quality of life, KCCQ=Kansas City Cardiomyopathy Questionnaire, LVEF=left ventricular ejection fraction, MI=myocardial infarction, MLWHFQ=Minnesota Living With Heart Failure questionnaire, NYHA=New York Heart Association, RCT=randomised controlled trial



To date, the only published study to have generated utility functions for HF patients that can be used in economic evaluations is based on a subsample of 1395 patients from the EPHEsus trial (Gohler et al., 2009). HRQoL was modelled with the EQ-5D at several time points using NYHA class and cardiovascular rehospitalisations as major independent variables, also adjusting for age, gender and cardiovascular comorbidities.

In addition, six further studies have derived utilities for relevant clinical subgroups, mostly by NYHA class (see Table 9). Three studies have used the direct elicitation techniques time-trade off and standard gamble, while four studies have used indirect techniques deriving utilities based on the EQ-5D, SF-6D and HUI-3. As the samples differ across studies, with settings ranging from primary and specialist outpatient care, inpatient care to even the general population, the different utilities by NYHA class are most likely a reflection of the underlying sample characteristics in terms of setting, age and LVEF. Interestingly, the general population sample yielded some of the highest utility estimates (Nichol et al., 2004).

In one of the first studies focusing on patient preferences for survival compared to improvements in symptoms, Lewis and colleagues (Lewis et al., 2001) used both the standard gamble and time trade-off techniques in a sample of advanced HF patients. They found a good correlation between the two preference measures. Physiologic measures of disease severity, specifically NYHA class, peak oxygen consumption and jugular venous pressure, were closely correlated with utility scores, while age, gender, duration of HF and LVEF were not. By collecting HRQoL information using the MLWHFQ, the study also highlighted the fact that HRQoL and utilities are two distinct concepts measuring overlapping, yet separate constructs. So although MLWHFQ scores and utilities showed modest correlation, the HRQoL scores were distributed evenly over the whole spectrum of possible answers, while patient preferences were strongly weighted towards the highest or lowest possible utility values (time-trade off scores). This suggests that the preferences reflect additional components apart from HRQoL or functional status that influence patients' willingness to trade all or none of their remaining life for achieving full health (Lewis et al., 2001).

More recently, Kontodimopoulos and colleagues (Kontodimopoulos et al., 2011) compared utilities derived from the EQ-5D and SF-6D in CHF patients undergoing elective cardiac surgery. Disease severity was measured with the Duke Activity Status Index rather than NYHA class. A strong correlation of EQ-5D and SF-6D values was seen, but agreement was poor in particular for subgroups with low functional capacity. The authors derived a cross-over point between EQ-5D and SF-6D utilities at 0.722, above which utility scores would be higher with the EQ-5D than the SF-6D, and below which the opposite would hold. Of note, the EQ-5D VAS was not able to differentiate between patient groups in terms of disease characteristics. In addition to the utilities by level of LVEF, the authors also reported utilities by other patient categories, e.g. gender (women lower than men) and age (no clear trend) (Kontodimopoulos et al., 2011).

**Table 9: Previous studies presenting utility weights for CHF**

Publication	Country	Setting	Sample size	Patient characteristics at baseline	Utility measure	Utility weights
<b>Lewis et al., 2001</b>	US	Hospital patients	99	Mean age 52 years Mean LVEF 24% NYHA class I-IV	TTO, SG	Estimates based on graph: NYHA I: 1.0 (TTO), 0.99 (SG) NYHA II: 0.9 (TTO), 0.8 (SG) NYHA III: 0.65 (TTO), 0.65 (SG) NYHA IV: 0.3 (TTO), 0.35 (SG)
<b>Nichol et al., 2004</b>	Canada	Convenience sample from general population	66	Mean age 49 years	SG	NYHA II: 0.82 NYHA III: 0.72 NYHA IV: 0.58
<b>Alehagen et al., 2008</b>	Sweden	Patients from primary care clinics	323	Mean age 72 years LVEF<30%: 1%, 30-39%: 9%, 40-49%: 14%, ≥50%: 76% Self-assessed NYHA class I-IV	TTO	Self-assessed NYHA NYHA I: 0.77 NYHA II: 0.68 NYHA III: 0.61 NYHA IV: 0.50 (incl. NYHA IIIb, i.e. not able to walk >200m)
<b>Miller et al., 2009</b>	US	Community sample, part of disease management trial	751	Mean age 70 years Mean LVEF 35% NYHA class I-IV	SF-6D	NYHA I: 0.75 NYHA II: 0.64 NYHA III and IV: 0.58
<b>Gohler et al., 2009</b>	International	RCT (EPHESUS) of eplerenone in post-acute MI patients; HRQoL sub-study	1395	Mean age 64 years Mean LVEF 32% NYHA class I-IV	EQ-5D	Univariate associations with utility: NYHA I: 0.855 NYHA II: 0.771 NYHA III: 0.673 NYHA IV: 0.532
<b>Pressler et al., 2011</b>	US	Convenience sample of CHF patients from primary care and heart clinics	210	Mean age 57 years NYHA class I-IV	HUI-3	NYHA I: 0.76 NYHA II: 0.56 NYHA III: 0.35 NYHA IV: 0.24
<b>Kontodimopoulos et al., 2011</b>	Greece	Hospital admissions for elective cardiac surgery	251	Mean age 66 years Mean LVEF 52%	EQ-5D, SF-6D	Utilities by level of LVEF: <50%: 0.621 (EQ-5D), 0.677 (SF-6D) 50-59%: 0.712 (EQ-5D), 0.702 (SF-6D) >60%: 0.755 (EQ-5D), 0.743 (SF-6D)

Notes: EPHESUS=Eplerenone Post-acute Myocardial Infarction Heart Failure Efficacy and Survival, HRQoL=health-related quality of life, HUI-3=Health Utilities Index Mark-3, LVEF=left ventricular ejection fraction, MI=myocardial infarction, NYHA=New York Heart Association, RCT=randomised controlled trial, SG=standard gamble, TTO=time trade-off

### **2.3.4 Impact of health-related quality of life on clinical outcomes in chronic heart failure**

Given the high mortality and morbidity in CHF, considerable research efforts have been invested to understand prognostic factors in this population. The role of HRQoL as an independent predictor of hospitalisations and mortality has been studied in different age groups, stages of CHF and settings, including primary care, specialist outpatient care, following hospitalisation and in randomised trials. Several disease-specific and generic HRQoL measurements have been used, including the MLWHFQ, KCCQ, SF-36 and health status questions. The follow-up period in most previous studies has been limited to between six months and five years. Only few studies with a follow-up of over five years have investigated HRQoL and mortality (Johansson et al., 2008; Lupon et al., 2013; Zuluaga et al., 2010), and there is a lack of long-term results concerning HRQoL and readmissions.

Previously identified predictors of mortality (all or cardiovascular causes) have included HRQoL, age, gender, NYHA class, disease history/duration, LVEF, comorbidities, clinical measures such as renal function and type of drug treatment (beta-blockers, angiotensin converting enzyme inhibitors, diuretics). Predictors of hospitalisations (all or cardiovascular causes) have included HRQoL, age, gender, NYHA class, comorbidities and beta-blockers.

Table 10 summarises studies relevant for the context of long-term prognosis, usually with sample sizes of at least 200 patients to allow for the high mortality rate (except for selected studies with HRQoL measures of specific interest, e.g. utilities based on the standard gamble technique or EQ-5D).

**Table 10: Selected studies on the role of HRQoL in HF prognosis with follow-up of more than one year**

Publication	Country	Setting	Sample size	Patient characteristics at baseline	HRQoL measure	Outcomes measure	Duration of follow-up	Main results
<b>Stull et al., 2001</b>	Inter-national	RCT of enalapril (SOLVD)	3884 (in multi-variate analyses)	Mean age 59 years LVEF $\leq 35\%$ , mean 27% NYHA class I-III	Battery of self-reported QoL questions (51 items)	HF-related hospitalisation	Mean 3 years; up to 5 years	HRQoL independent predictor in multivariate analyses for all age groups, with different clinical predictors
<b>Soto et al., 2004</b>	Inter-national	RCT (EPHESUS) of eplerenone in post-acute MI patients; HRQoL sub-study	1516	Mean age 64 years Mean LVEF 33% NYHA class I-IV	KCCQ	Combined endpoint of 1-year cardiovascular death or hospitalisation	Mean 16 months	Independent predictors combined endpoint: age, COPD, diabetes, prior MI, prior HF, post-MI LVEF, beta-blockers, KCCQ summary score
<b>Mejhert et al., 2006</b>	Sweden	Patients hospitalised for HF	208	Mean age 76 years Mean LVEF 34% NYHA class II-IV	NHP	All-cause mortality and rehospitalisation	Mean 3.1 years	Rehospitalisation: NHP total score only significant predictor; Mortality: age, gender, BNP, creatinine, LV systolic function independent predictors
<b>Sullivan et al., 2007</b>	US	Heart failure/ pre-transplant clinic	142	Mean age 53 years Mean LVEF 27% Mean NYHA class 2.7	SF-36 (health status question), KCCQ summary score, feeling thermometer, SG	Combined endpoint of heart transplant or death	Mean 3 years	Standard gamble only significant health status measure when controlling for significant clinical variables
<b>Faller et al., 2007</b>	Germany	Outpatients at university centres	231	Mean age 64 years Mean EF 44% NYHA class I-IV	SF-36, KCCQ, Patient Health Questionnaire (depression)	All-cause mortality	Median 2.7 years	Significant predictors in multivariate models: SF-36 mental score - LVEF, NYHA class; KCCQ - age, NYHA class
<b>Subramanian et al., 2007</b>	US	Veteran Affairs outpatients	494	Mean age 68 years Left ventricular dysfunction NYHA class I-IV	SF-36, KCCQ, McMaster Chronic Heart failure Questionnaire	All-cause mortality	5 years	Independent predictors: age, COPD, hypertension, hyperlipidaemia, other comorbidities, number of previous hospitalisations, beta-blockers, diuretics, creatinine, sodium, heart rate, KCCQ change over time score
<b>Johansson et al., 2008</b>	Sweden	Patients treated in primary care	448	Mean age 73 years 12% with LVEF $<40\%$ NYHA class I-III	Health status question from SF-36	All-cause and cardiovascular mortality	10 years	Independent predictors cardiovascular mortality: global health perception at baseline, age, gender, NYHA, diabetes, LVEF, BNP

<b>Iqbal et al., 2010</b>	UK	HF patients discharged from cardiology wards or outpatient clinics	179	Mean age 71 years NYHA class I-IV	EQ-5D, MLWHFQ	All-cause mortality, all-cause and HF rehospitalisation	Mean follow-up 3.3 years	Survival only predicted by previous HF hospitalisation and lack of beta-blockers in multivariate analyses; HF rehospitalisation by MLWHFQ score and beta-blockers
<b>O'Loughlin et al., 2010</b>	Ireland	Survivors of NYHA class IV hospital admission enrolled in disease management programme	225	Mean age 69 years 78% had LVEF<45% 87% had NYHA class I-II	MLWHFQ	All-cause mortality, all-cause emergency admission	Mean 18 months	Predictors multivariate models: Mortality - HRQoL, HRQoL*age interaction, age, gender, creatinine, systolic blood pressure; Emergency admission - HRQoL, HRQoL*age, age, gender, NYHA class, cancer
<b>Zuluaga et al., 2010</b>	Spain	Patients admitted for emergency hospitalisation	416	Mean age 77 years 45% had LVEF≤45% 34% had NYHA class III-IV	SF-36, MLWHFQ	All-cause mortality	7 years	Independent predictors: HRQoL, age, ischaemic aetiology, NYHA class, comorbidity index, COPD, creatinine, ACEI, beta-blockers, recent HF hospitalisation
<b>Chamberlain et al., 2013</b>	US	Community sample covering in- and outpatients, HFPEF and HFREF	352	Mean age 73 years	SF-12, physical component score, self-rated general health	All-cause mortality	Mean 2.3 years	Self-rated health and functional exercise capacity independently predicted mortality in adjusted Cox models
<b>Hoekstra et al., 2013</b>	The Netherlands	Patients hospitalised for HF	661	Mean age 71 years Mean LVEF 34% NYHA class II-IV	Ladder of life, RAND-36, MLWHFQ	All-cause mortality	3 years	In model adjusted for demographic and clinical variables, only physical functioning and general health dimensions of RAND36 significant
<b>Lupon et al., 2013</b>	Spain	Outpatients referred to HF university hospital clinic	1151	Median age 69 years Median LVEF 31% NYHA class I-IV	MLWHFQ	All-cause mortality	Up to 6 years	Predictors multivariate model: HRQoL, age, gender, HF duration, NYHA class, diabetes, anaemia, peripheral vascular disease, beta-blockers, ACEI/ARB

Notes: ACEI=angiotensin-converting-enzyme inhibitor, ARB= angiotensin II receptor blocker, BMI=body mass index, BNP=brain natriuretic peptide, COPD=chronic obstructive pulmonary disease, eGFR=estimated glomerular filtration rate, EPHEUS=Eplerenone Post-acute Myocardial Infarction Heart Failure Efficacy and Survival, HFPEF=heart failure with preserved ejection fraction, HFREF=heart failure with reduced ejection fraction, HRQoL=health-related quality of life, KCCQ=Kansas City Cardiomyopathy Questionnaire, LVEF=left ventricular ejection fraction, MI=myocardial infarction, MLWHFQ=Minnesota Living With Heart Failure Questionnaire, NHP=Nottingham Health Profile, NYHA=New York Heart Association, SG=standard gamble, SOLVD= Studies of Left Ventricular Dysfunction

### **2.3.5 Identified knowledge gaps**

Based on the focused literature review on the different subtopics, knowledge gaps could be identified in the following areas:

- Utilities in AF and driving factors based on a comprehensive sample of patients covering relevant clinical variables, including AF type (first detected, paroxysmal, persistent and permanent) and adverse events;
- Cost-effectiveness analyses evaluating AF treatments using utilities from a large comprehensive patient sample;
- Utilities in CHF and driving factors based on a comprehensive sample of patients from clinical practice covering different treatment settings and levels of LVEF;
- Prognostic value of HRQoL in CHF for death over a timeframe of more than six years and for rehospitalisations over more than three years.

### **3 AIMS OF THE THESIS**

The overarching aim of this thesis is to enhance the knowledge on what drives HRQoL and utilities in patients with AF and CHF, and how these in turn influence clinical and economic outcomes. Specifically, the following four research questions have been investigated in the included papers:

- What factors drive utility in patients with AF at different time points? (Paper I)
- What is the cost-effectiveness of a new treatment for AF, given data from a morbidity/mortality trial and derived utility weights? (Paper II)
- What factors drive utility in patients with CHF and what statistical model is most suited for the data? (Paper III)
- What is the impact of HRQoL on long-term morbidity and mortality in patients hospitalised with systolic CHF? (Paper IV)





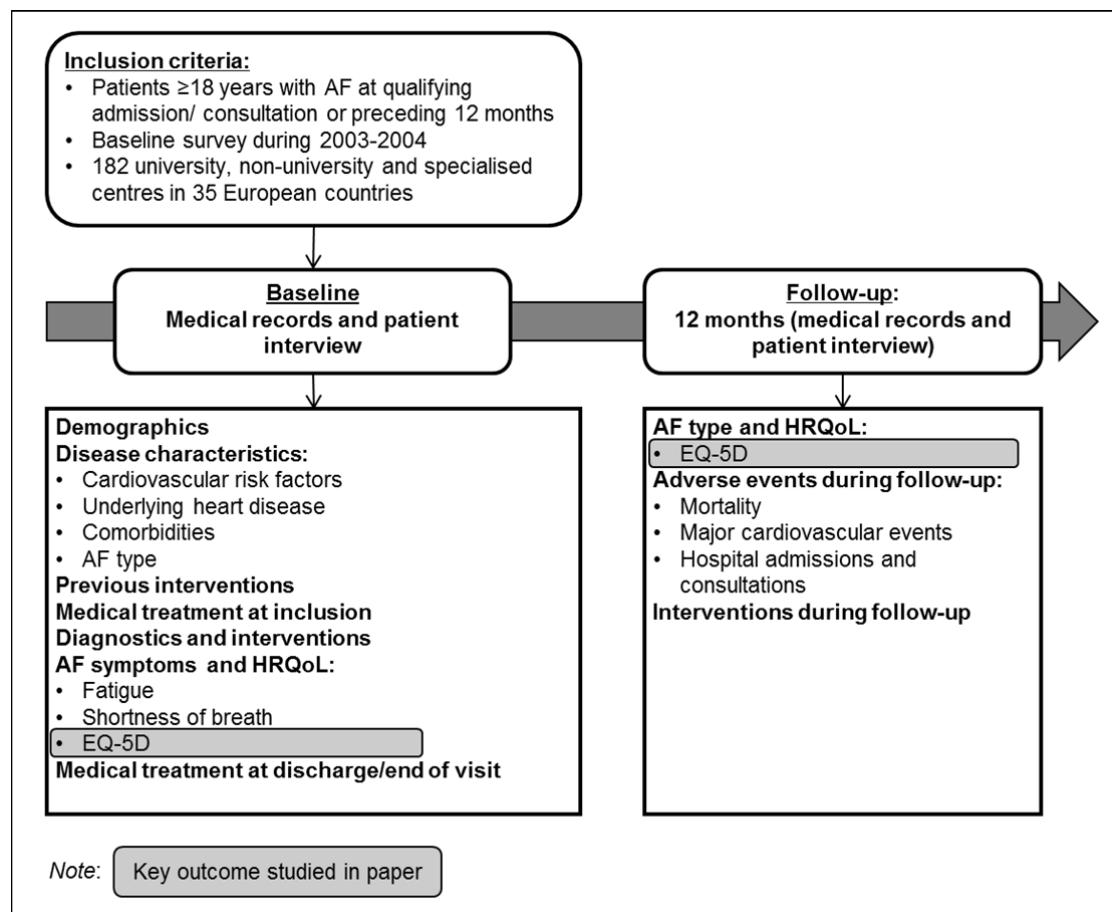
## 4 MATERIALS AND METHODS

### 4.1 MATERIALS

#### 4.1.1 The Euro Heart Survey on Atrial Fibrillation

The Euro Heart Survey on AF was an observational study enrolling 5333 patients from cardiology clinics across 35 European countries during 2003 and 2004 (Nieuwlaat et al., 2005). Inclusion criteria were age  $\geq 18$  years and AF on electrocardiogram or Holter recording at the enrolment visit or during the preceding 12 months. The overarching objective of the study was to describe and compare clinical practice against guidelines. A range of demographic, clinical, treatment and functional variables were collected at baseline, including the EQ-5D for measuring HRQoL. After one year, information was collected on AF type, HRQoL, adverse events and interventions during follow-up. Figure 11 summarises information about the patient sample, key variables and outcomes studied in Paper I.

**Figure 11: Study flow chart describing patient sample, key variables and outcomes of interest – Paper I (Euro Heart Survey on Atrial Fibrillation)**



Notes: AF=atrial fibrillation, HRQoL=health-related quality of life

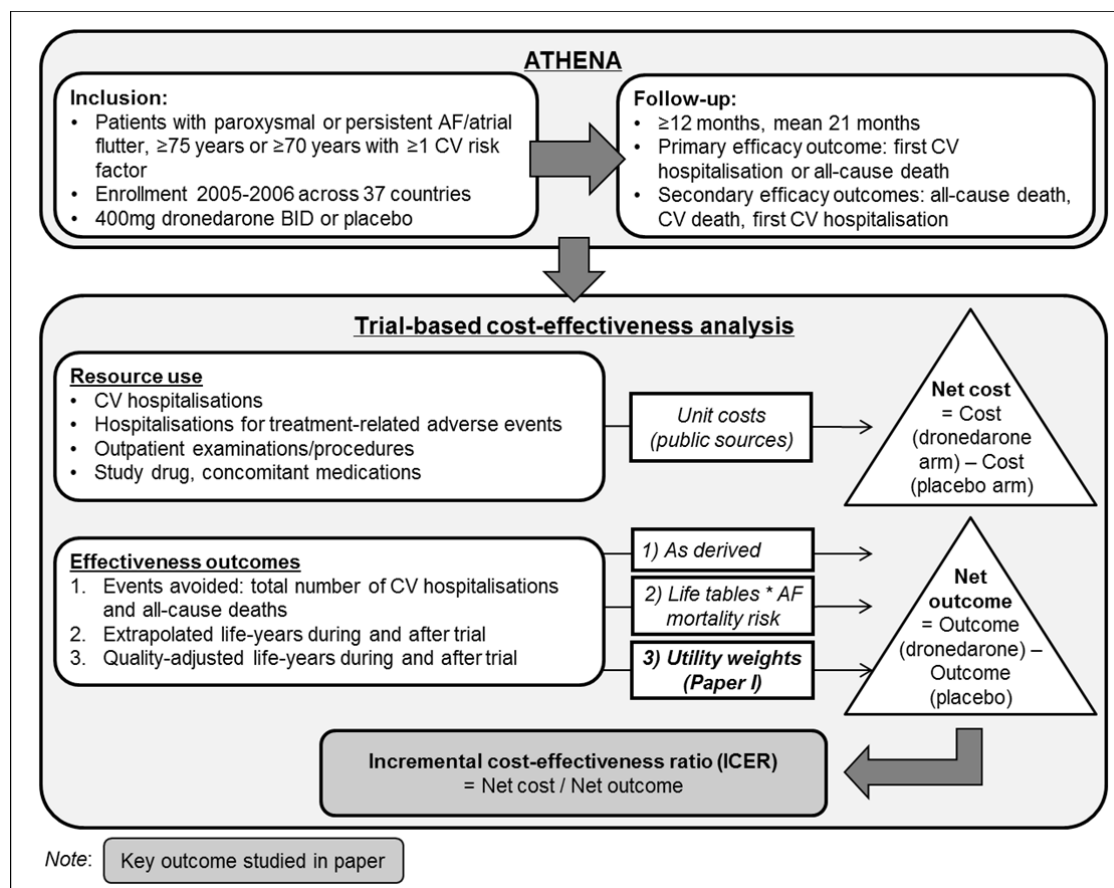
At baseline, 5050 patients (corresponding to 95%) had complete information on the EQ-5D, compared to 3045 (76% of surviving patients not lost to follow-up) at one year. Mean age at

baseline was 66 years, 58% were male. There was a relatively even distribution amongst AF types (first detected, paroxysmal, persistent and permanent); 70% of patients were experiencing symptoms. At follow-up, only 34% of patients were symptomatic; at the same time, there had been a shift towards permanent AF representing the most frequent AF type (41%).

#### 4.1.2 The ATHENA trial

ATHENA (A placebo-controlled, double-blind, parallel arm Trial to assess the efficacy of dronedarone 400mg bid for the prevention of cardiovascular Hospitalization or death from any cause in patiENts with Atrial fibrillation/atrial flutter) was the first clinical trial in AF to incorporate morbidity and mortality outcomes in its primary endpoint. It evaluated the effect of dronedarone, a new antiarrhythmic drug, against placebo on top of standard of care in patients with paroxysmal or persistent AF or atrial flutter, who were  $\geq 75$  years or  $\geq 70$  years with at least one cardiovascular risk factor. The study enrolled 4628 patients in 37 countries, who were followed for at least 12 months (Hohnloser et al., 2009). Figure 12 summarises information about the patient sample, key variables and outcomes studied in Paper II. The methods relevant to the trial-based cost-effectiveness analysis are detailed in Section 4.2.2.

**Figure 12: Study flow chart describing patient sample, key variables and outcomes of interest – Paper II (ATHENA within-trial analysis)**



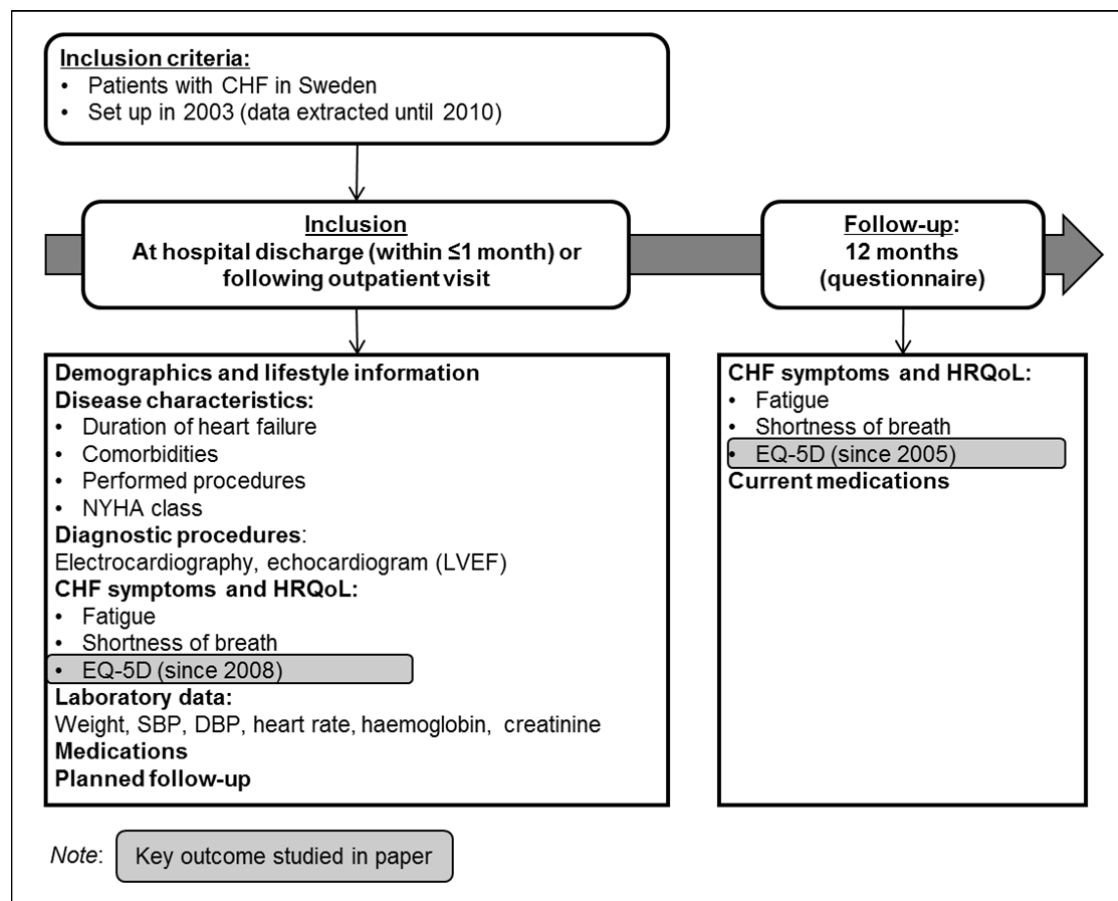
Notes: AF=atrial fibrillation, BID=two times a day, CV=cardiovascular

At baseline, mean age in the trial population was 72 years. In the dronedarone arm, 51% were male, compared to 55% in the placebo arm. Hypertension was the most common comorbidity (86%), 60% had structural heart disease and 21% a history of CHF (Hohnloser et al., 2009).

#### 4.1.3 The Swedish Heart Failure Registry

The Swedish Heart Failure registry was established in 2003 to describe CHF management in Swedish clinical practice, with the ultimate goal of improving the quality of care. Inclusion can occur at discharge from hospital (within one month) or following an outpatient visit. The majority of patients have been included from hospitals (93% of all 50 827 patients in 2012), with the rest stemming from primary care (Vasko, Jonsson & Dahlstrom, RiksSvikt, 2013). Information is collected on demographics, lifestyle factors, disease characteristics, diagnostics, laboratory measures, medications, CHF symptoms and HRQoL; the latter is measured with the EQ-5D. Follow-up occurs after one year through a questionnaire covering functional capacity, HRQoL and current medications. Figure 13 summarises information about the patient sample, key variables and outcomes studied in Paper III.

**Figure 13: Study flow chart describing patient sample, key variables and outcomes of interest – Paper III (Swedish Heart Failure Registry)**



Notes: CHF=chronic heart failure, DBP=diastolic blood pressure, HRQoL=health-related quality of life, LVEF=left ventricular ejection fraction, NYHA=New York Heart Association, SBP=systolic blood pressure

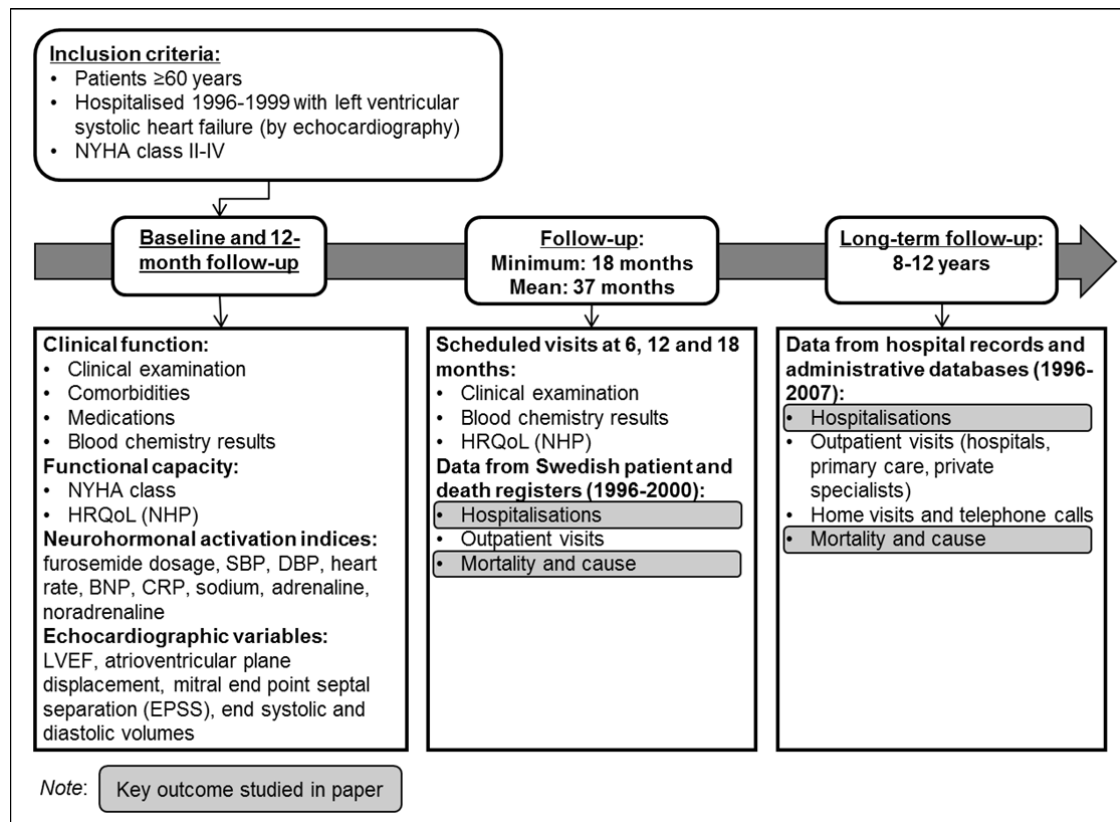
In the sample extracted until the end of 2010, 5334 patients had completed the EQ-5D at inclusion (14% of all patients), compared to 3495 of patients at 1-year follow-up (66% of those with EQ-5D at baseline). Mean age at inclusion was 73 years and 65% were male. The majority of patients were in NYHA classes II-III, and 17% had HFPEF. The low proportion of patients with complete HRQoL information at baseline is partly due to this information only being collected since 2008, and partly a reflection of a higher priority given to clinical information at the time of inclusion (which often coincides with a hospitalisation).

#### **4.1.4 The OPTIMAL study**

The Optimising congestive heart failure outpatient clinic project (OPTIMAL) was a randomised prospective study in patients  $\geq 60$  years hospitalised with systolic HF and NYHA class II-IV at Danderyd University Hospital in Stockholm, Sweden, during 1996-1999. The primary study endpoint was the impact of a nurse monitored management programme on HRQoL. A range of clinical parameters, echocardiographic variables and neurohormonal activation indices were collected at baseline, together with information on functional capacity and HRQoL, which was measured using the generic NHP instrument. Follow-up occurred through scheduled visits at 6, 12 and 18 months; long-term follow-up over 8-12 years was performed through medical records and administrative databases. Figure 14 summarises information about the patient sample, key variables and outcomes studied in Paper IV.

Of the 208 patients included in the study, 177 had completed the NHP at baseline. Mean age was 76 years and 56% were male. The vast majority were in NYHA classes II-III, mean LVEF was 34%.

**Figure 14: Study flow chart describing patient sample, key variables and outcomes of interest – Paper IV (OPTIMAL)**



Notes: BNP=brain natriuretic peptide, CRP=C-reactive protein, DBP=diastolic blood pressure, HRQoL=health-related quality of life, LVEF=left ventricular ejection fraction, NHP=Nottingham Health Profile, NYHA=New York Heart Association, SBP=systolic blood pressure

## 4.2 METHODS

### 4.2.1 Regression analysis

#### 4.2.1.1 Modelling utility

In Papers I and III, the key outcome variable of interest was utility derived from the EQ-5D; in Paper III, change in utility over one year was also analysed. Utilities are usually not normally distributed, as they tend to have a “spike” at the ceiling value, e.g. 1 for the EQ-5D when using the health state valuations from the UK general public. Moreover, they are bounded at the possible maximum and minimum values that can be obtained using the respective algorithm for converting health states to utilities. Therefore, linear regression using ordinary least squares (OLS) may not provide correct estimations for the significance of model parameters, as two important assumptions underlying OLS regression do not hold, namely that the outcome variable should be normally distributed and that the error terms should be normally distributed with mean zero and constant variance (homoscedasticity).

A number of different techniques have been proposed and tested to address the statistical properties of utility data, yet there is no single method that has been shown to be generally

preferable to date. At the time of designing and conducting the analyses for Paper I, a method that had generated interest in the context of utility data was Powell's censored least absolute deviations (CLAD) estimator (Powell, 1984). Its benefits are that it does not depend on distributional or homoscedasticity assumptions and is robust to censoring. However, this is achieved by modelling the median rather than the mean, which is a drawback for economic evaluations that inherently focus on mean values. In Paper I, we therefore used the CLAD method for deriving covariates that significantly impacted on utility through stepwise regression, and then used these variables in an OLS regression to obtain mean coefficients for prediction.

The fact that CLAD (and Tobit) regression assumes that utilities are truncated at 1, i.e. that it conceptually would be possible to obtain health states that are better than perfect health, has subsequently been criticised from a theoretical standpoint and has also been shown to lead to bias in the mean estimates. Instead, OLS regression using robust standard errors or non-parametric bootstrapping has been recommended as an unbiased estimator producing valid confidence intervals in sufficiently large samples (Pullenayegum et al., 2010). An important caveat is that the OLS model needs to be correctly specified, e.g. taking interaction terms into account and using suitable functional forms for model covariates, in particular checking for non-linearity (Kirkwood & Sterne, Blackwell Science, 2005; Pullenayegum, Wong & Childs, 2013). Therefore, we chose to use OLS with robust standard errors as the main analytical approach in Paper III. The model selection was based on initial univariate and subsequent multivariate regressions. In addition, we performed a sensitivity analysis using two-part models, which have been proposed as an alternative method for utility data, as they are particularly suited for modelling the ceiling effect; a disadvantage is that they do not provide outputs that can be directly used for the calculation of means (Huang et al., 2008; Pullenayegum et al., 2010). We compared the model predictions for the two regression methods using within-sample validation, comparing accuracy on an aggregate and individual level as well as prediction ranges.

For the analyses relating to change in utility over one year, this outcome variable is not associated with the same distributional issues as utility. In our samples for Paper III, the change in utility had a high kurtosis, but was otherwise relatively normally distributed. Therefore, we used OLS with robust standard errors to determine drivers of change.

#### *4.2.1.2 Time-to-event analyses*

In Paper IV, the key outcome variables were time to death and time to rehospitalisation, for all or cardiovascular causes. We used Cox proportional hazard regression to model the relationship between demographic, clinical and functional variables and the respective outcome of interest. Since the NHP instrument does not have a defined summary score, exploratory factor analysis was used to determine the groups (factors) consisting of HRQoL dimensions that explained most of the common variance (Fayers & Machin, John Wiley & Sons Ltd., 2007); as there is weak correlation between the derived factors, these could then be used as covariates in the Cox regression models. The proportional hazards assumption of the

Cox model, i.e. that the ratio of the hazards between different patient groups remains constant over time, was tested using Schoenfeld residuals.

#### **4.2.2 Cost-effectiveness analysis**

In Paper II, we performed a trial-based cost-effectiveness analysis of dronedarone vs. placebo on top of standard of care, using patient-level data from ATHENA. The analysis was based on a Canadian payer perspective, and different outcomes were assessed over the trial duration and extended to a lifetime horizon. We tested for heterogeneity across countries and differences in baseline utility. Sensitivity analyses were performed around different input parameters, including scenario and subgroup analyses to mirror label restrictions for dronedarone following later results from the PALLAS (Permanent Atrial Fibrillation Outcome Study Using Dronedarone on Top of Standard Therapy) trial. Uncertainty was presented using confidence intervals and cost-effectiveness acceptability curves for the different outcomes.

Treatment-related resources were based on the case report forms, excluding protocol-driven costs such as regular laboratory tests. Thus, the types of resources covered the following: cardiovascular hospitalisations, hospitalisations due to treatment-related adverse events, outpatient examinations and procedures, study drug and concomitant medications. Costing was performed using standard public sources. In line with Canadian guidelines, we used a health care payer perspective, which does not cover costs due to lost productivity or other non-medical resources such as home help. Considering the relatively high mean age of patients with AF, both in general and particularly in ATHENA, costs due to lost productivity contribute with a comparatively small proportion to overall costs. For example, in an analysis based on the Euro Heart Survey on AF, work loss was found to account for 4-17% of total annual costs across the five countries contributing most patients (Ringborg et al., 2008). Mean age in these countries was 68 years, compared to 72 years in ATHENA, where 19% of patients were younger than 65 years (Hohnloser et al., 2009). In this light, the impact of perspective on total costs is smaller than for diseases affecting younger populations. Moreover, for dronedarone, the reduction of cardiovascular events and mortality seen with treatment could in a societal perspective also be expected to be reflected in reduced sick leave and reduced need for services outside the health care system.

We evaluated three different outcome measures:

- Number of events avoided: based on the primary efficacy endpoint in ATHENA of cardiovascular hospitalisations or all-cause death during the trial period;
- Number of life years gained during and after the clinical trial: based on predicted survival for patients alive at the end of the trial, using life tables by age and gender adjusted for an AF-specific mortality risk;
- Number of QALYs gained during and after the clinical trial: based on predicted survival weighted with utilities derived from the regression analyses of follow-up data in Paper I.

Each of the outcome measures is related to some methodological issues that need to be taken into account. An outcome such as events avoided is very specific to the data at hand, and does not allow comparison with results even for other AF interventions assessed using other outcomes, e.g. time to AF recurrence. Moreover, as the event in this case included cardiovascular hospitalisations, which were also an important cost item in the analysis, there is a risk of double-counting its impact due to inclusion as part of both costs and effects.

The extrapolation of survival beyond the trial timeframe for the estimation of life years gained is linked to certain assumptions. Often, the mortality observed in clinical trials is lower than can be expected in natural history cohorts, due to selected patient populations. In this case, use of a parametric regression model to predict survival after the end of the trial would have overestimated the number of remaining life years. Therefore, life tables were used instead, which do not require any distributional assumptions. These were multiplied by an AF-specific mortality risk derived from a similar patient population. No further benefit of treatment was assumed after the trial (when treatment was assumed to stop). In the absence of follow-up data, this can generally be seen as an acceptable assumption. The only exception would be if treatment discontinuation were related to some kind of rebound effect, i.e. patients discontinuing treatment in the active arm were to progress faster than those in the comparison group. If that were the case, the benefit of treatment would have been overestimated.

As neither HRQoL nor utilities were collected as part of ATHENA, QALYs could only be estimated through the use of external data sources. The results from the Euro Heart Survey in AF allowed estimation of utilities based on adjustment for known covariates in patient groups covered by both studies. As the underlying samples were different e.g. with regards to AF type (permanent AF was not part of ATHENA), there could potentially remain a risk of underestimation of utilities when applying the regressions based on the Euro Heart Survey.



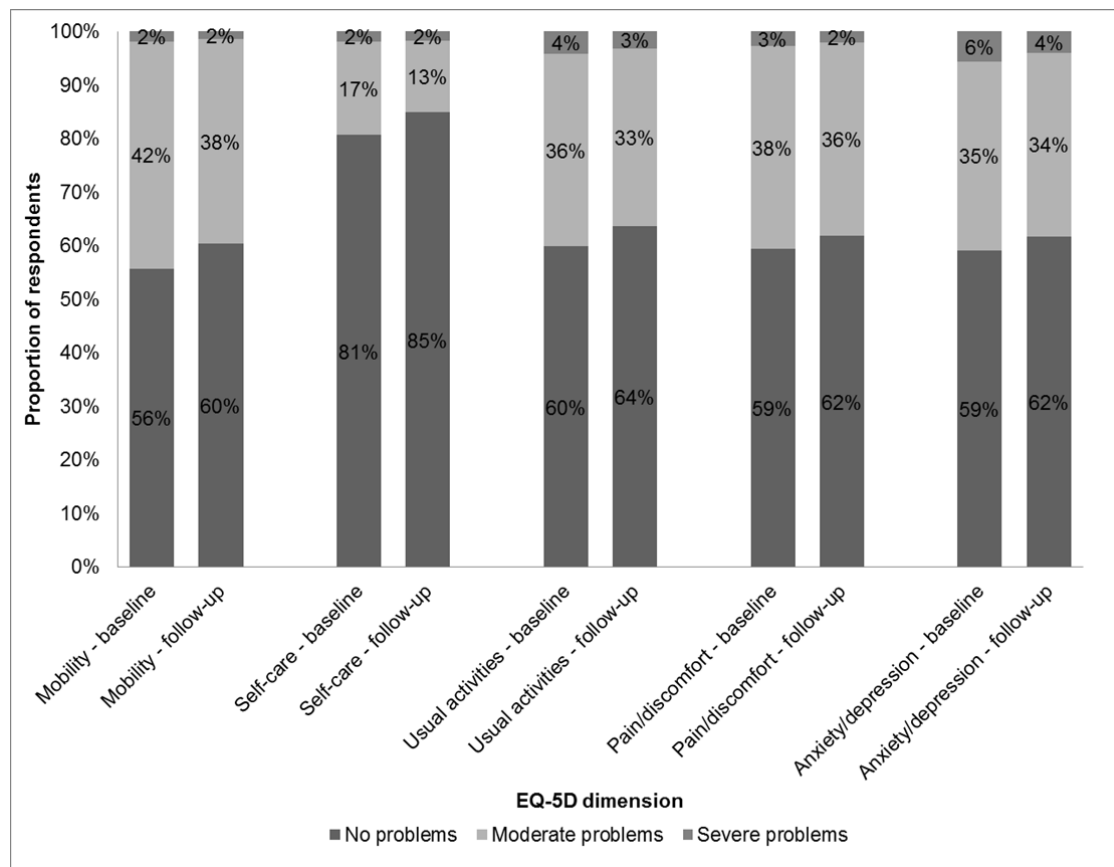
## 5 RESULTS

### 5.1 DETERMINANTS OF UTILITY IN ATRIAL FIBRILLATION (PAPER I)

At baseline, mean utility was 0.751 (SD 0.269); at follow-up this was slightly higher at 0.779 (SD 0.253). Amongst those who completed the EQ-5D at both time points, the average improvement in utility was 0.013 (SD 0.261). Patients with permanent AF had the lowest utility regardless of measurement occasion.

The dimension of mobility was most affected in the sample, with 44% and 40% reporting some or severe problems at baseline and follow-up, respectively. Self-care was least affected, with 15-19% reporting some or severe problems over time. The distribution of responses across severity levels for the EQ-5D descriptive system is shown in Figure 15. The overall severity of problems had lessened at 1-year follow-up, which is likely a reflection of the measurement setting.

**Figure 15: Distribution of responses by severity levels in the EQ-5D dimensions for AF patients at baseline and 1-year follow-up (adapted from Berg et al., 2010)**



At baseline, increasing age, female gender, domestic status outside the own home (e.g. nursing homes or homes for the elderly), existing comorbidities (including diabetes, previous stroke), as well as symptoms of AF, CHF or angina were all associated with reduced utility. Amongst these, previous stroke and CHF symptoms had the strongest impact. Regular physical activity had a positive effect on utilities. Utilities also differed by AF type, with

paroxysmal and permanent AF having significantly lower utilities than those with a first detected episode (reference group).

At follow-up, major adverse events during the follow-up period, specifically stroke, myocardial infarction and CHF, led to large decrements in utility. Presence of AF symptoms at follow-up was also a significant negative driver of utility. Similar patterns as at baseline were seen for some of the variables, e.g. age and gender. To account for the important role of utility at baseline for development over time, determinants of baseline utility were included in the follow-up model regardless of statistical significance.

## 5.2 ROLE OF DIFFERENT OUTCOME MEASURES IN COST-EFFECTIVENESS ANALYSES OF A NEW TREATMENT FOR ATRIAL FIBRILLATION (PAPER II)

During the ATHENA trial, there were on average 0.51 cardiovascular hospitalisations per patient in the dronedarone arm, compared to 0.69 in the placebo arm. Including all primary endpoint events (i.e. cardiovascular hospitalisations and death from any cause), there were 0.18 fewer events in the dronedarone arm. When extrapolating survival beyond the trial period using AF-adjusted life tables (with a hazard ratio of 1.66), expected survival was 0.22 life years higher for patients with dronedarone, corresponding to 0.13 QALYs. Table 16 summarises the mean effects for the different outcome measures per trial arm, showing both undiscounted and discounted results.

**Table 16: Mean effects per patient during the trial period (95% CI) (Berg et al., 2013)**

	Dronedarone	Placebo
<b>Undiscounted</b>		
Cardiovascular hospitalisations (number of events)	0.51 (0.47-0.55)	0.69 (0.64-0.74)
Events (CV hospitalisations and death from any cause)	0.56 (0.52-0.61)	0.75 (0.69-0.80)
Expected survival (years)	10.89 (10.66-11.13)	10.67 (10.44-10.91)
Expected quality-adjusted survival (QALYs)	8.62 (8.42-8.82)	8.48 (8.28-8.69)
<b>Discounted 5%</b>		
Expected survival (years)	8.21 (8.08-8.34)	8.08 (7.95-8.21)
Expected quality-adjusted survival (QALYs)	6.50 (6.39-6.61)	6.41 (6.29-6.53)

*Notes: CI=confidence interval, CV=cardiovascular, QALY=quality-adjusted life year*

Overall, mean undiscounted costs per patient were CAD\$694 higher in the dronedarone than placebo arm over the trial period (CAD\$7402 vs. CAD\$6708, respectively), which was due to higher treatment costs. These were to some extent offset by savings for cardiovascular hospitalisations, examinations and procedures, and other concomitant medications. The resulting base case ICERs were the following for the different outcomes:

- Using no discounting: CAD\$3784 per event avoided, CAD\$3141 per life year gained and CAD\$5190 per QALY;
- Using 5% discount rate: CAD\$3807 per event avoided, CAD\$5204 per life year gained and CAD\$7560 per QALY.

Amongst the different scenarios tested in sensitivity analyses, those relating to changes in assumptions and sources for survival extrapolation or utilities had limited impact on overall results. The largest impact was seen when using the lowest and highest costs per case mix group, with dronedarone treatment becoming the dominant strategy when the highest costs were used.

### **5.3 DETERMINANTS OF UTILITY IN CHRONIC HEART FAILURE (PAPER III)**

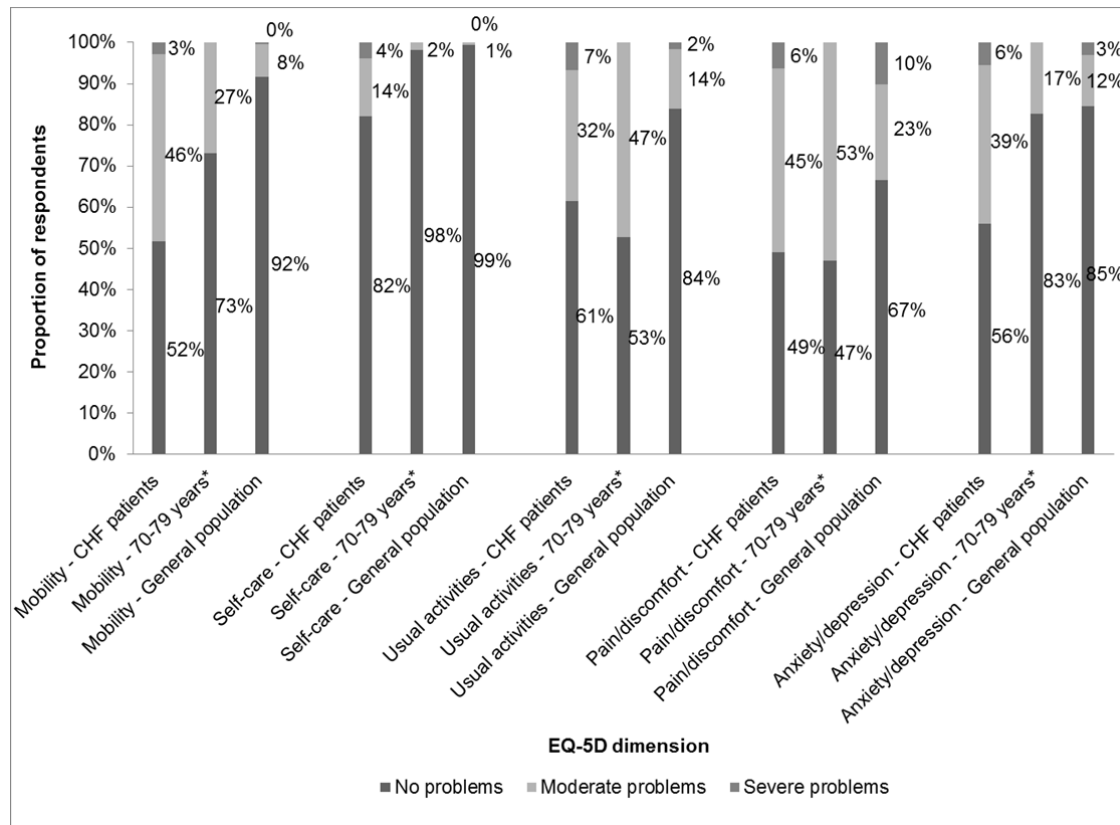
Amongst the sampled CHF patients, the dimensions for which the largest proportion of patients reported some or severe problems were mobility (48%), pain/discomfort (51%) and anxiety/depression (44%). Mobility, self-care and anxiety/depression were negatively affected when comparing to a similar age group (70-79 years) in the Swedish population (Burstrom, Johannesson & Diderichsen, 2001), whilst slightly fewer CHF patients reported some or severe problems with usual activities compared to the general population (Figure 17).

Utility at baseline was negatively affected by female gender, increasing age, increasing NYHA class, comorbidities (including lung disease and diabetes), and use of nitrates, antiplatelets or diuretics. For LVEF, those with HFPEF had a significantly lower utility than those with HFREF. Higher systolic blood pressure and haemoglobin levels were associated with increased utility, as were use of angiotensin converting enzyme inhibitors/angiotensin receptor blockers and beta-blockers. NYHA class was the single most influential determinant. We also identified a significant interaction between age category and NYHA class, whereby patients in the youngest age group were more severely affected by worse NYHA status than older patients.

In this data set, the OLS model performed slightly better than the two-part model on a population level and for capturing utility ranges. There was a tendency for the two-part model to generate somewhat more precise estimates on the individual level.

There was a noticeable loss to follow-up over one year (34% amongst those with information on the EQ-5D), which is a reflection of the high mortality and disease progression seen with CHF. Within the sample who filled in the EQ-5D at both time points, absolute change in utility over time was driven by gender, age and, measured at inclusion, NYHA class, disease duration, systolic blood pressure, ischaemic heart disease, lung disease, angiotensin converting enzyme inhibitors/angiotensin renin blockers and antiplatelets.

**Figure 17: Distribution of responses by severity levels in the EQ-5D dimensions for CHF patients at inclusion, the general Swedish population (16-84 years) and comparable age group (70-79 years)**



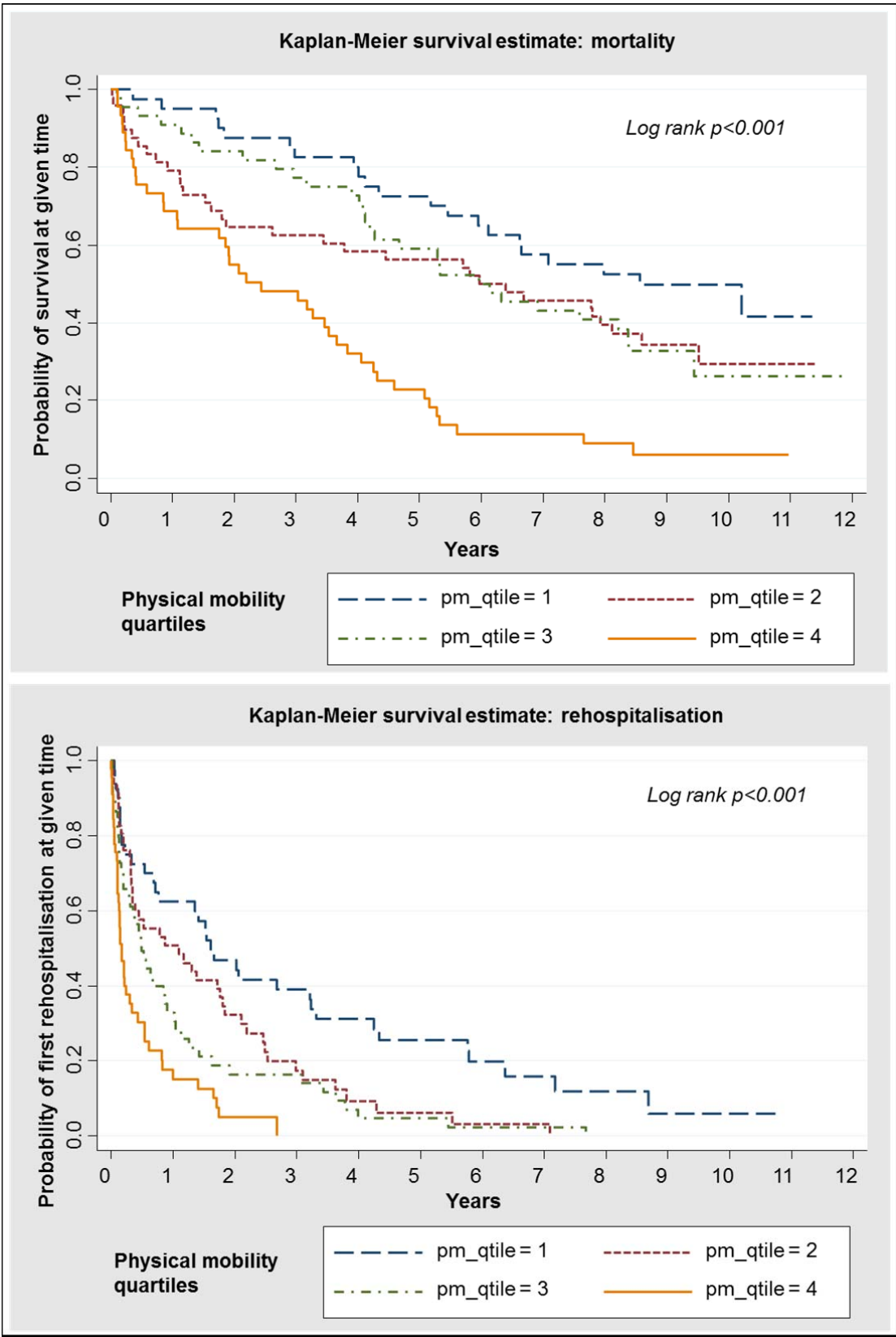
Notes: \*Only moderate or severe problems reported for 70-79 year olds.

## 5.4 IMPACT OF HEALTH-RELATED QUALITY OF LIFE ON PROGNOSIS IN CHRONIC HEART FAILURE (PAPER IV)

Physical mobility and emotional reactions were identified as the two NHP dimensions that contributed most information to the factor model, and were thus used in subsequent analyses. There was a clear and significant relationship between different strata of the physical activity dimension and time to death and time to rehospitalisation, as illustrated in Figure 18. This was also confirmed as part of the multivariate survival models, where physical mobility was a significant independent predictor for all four outcomes, with every 1% worsening (increase) resulting in a 1-2% increase in the hazard ratio of being hospitalised or dying. Emotional reactions were an additional independent predictor for all-cause hospitalisations, leading to a 1.3% increase in the hazard ratio with every 1% worsening.

As for previous studies in this field, it should be noted that the prognostic impact of HRQoL on outcomes has been shown in a statistical sense. The underlying causal mechanism between physical mobility and outcomes in CHF may be explained by the strong link to oxygen uptake and exercise capacity, which have previously been shown to be predictors of mortality (Mejhert et al., 2002). The mechanisms around other HRQoL domains, such as emotional reactions, and rehospitalisations warrant further research.

**Figure 18: Kaplan-Meier curves for all-cause mortality and hospitalisations by quartiles of NHP physical mobility score (Berg et al., 2014)**



### 5.4.1 Previously unpublished results: mortality impact of health-related quality of life measured at 1-year follow-up visit in outpatient care

As part of the OPTIMAL study, clinical and functional variables were collected at scheduled outpatient visits 6, 12 and 18 months after the qualifying admission. We hypothesised that there may also be an impact of HRQoL on prognosis when measured in an outpatient setting following the acute phase. To test this, we applied the same analytical approach as in Paper IV to the data collected as part of the 12-month follow-up. We chose this timeframe to allow for completed titration of medications and clinical stabilisation following the index hospitalisation. As the majority of first readmissions occurred during the first year of follow-up, we only focused on mortality in these analyses.

At 12-month follow-up, 150 patients had completed the NHP. Mean age was 76.2 years (SD 6.9), and 59% were male. Similar to baseline, physical mobility and emotional reactions were the most informative factors of the NHP. The prognostic value of HRQoL remained significant for cardiovascular mortality; in fact, physical mobility measured at 12 months in outpatient care had almost the same magnitude of effect as seen for the baseline measurement (Table 19). No significant impact of HRQoL on all-cause mortality was found using 12-month data, which could be due to the small sample size, but also due to competing risks from other events as patients survive the first year following the index hospitalisation. These results suggest that the impact of HRQoL, and specifically physical mobility, on cardiovascular mortality may be similar regardless of timing and setting of assessment, which provides useful information for clinical applications.

**Table 19: Final model specifications for predictors of cardiovascular mortality, using measurements from 12-month follow-up visit (n=112)**

Covariate	Hazard ratio	95% CI		p
Age at follow-up (years)	0.991	0.922	1.064	0.794
Female gender	0.445	0.216	0.917	0.028
eGFR at follow-up (ml/min)	0.974	0.952	0.997	0.029
BNP at follow-up (ng/l)*	3.796	1.846	7.803	0.000
Physical mobility at follow-up	1.013	1.000	1.026	0.041

Notes: \*Log-transformed variable

BNP=brain natriuretic peptide, eGRF=estimated glomerular filtration rate

## 6 DISCUSSION

### 6.1 CRITICAL EVALUATION OF STUDY METHODS AND FINDINGS

Table 20 summarises the key strengths and limitations of Papers I-IV, together with information on how limitations were sought to be addressed and, where relevant, the potential impact on results.

**Table 20: Summary of key strengths and limitations of Papers I-IV**

Study	Strengths	Limitations
<b>Paper I: Utilities in AF</b>	<ol style="list-style-type: none"> <li>1. Large international sample covering broad spectrum of AF (n=5050) – previously more narrow samples</li> <li>2. EQ-5D-based utilities useful for economic evaluations</li> <li>3. Use of statistical methods recommended for utility data at the time (CLAD), comparison with standard OLS models</li> </ol>	<ol style="list-style-type: none"> <li>1. Large share of specialised or interested centres, i.e. more severely ill patients</li> <li>2. Patients responding at baseline more healthy, i.e. potential underestimation of some effects</li> <li>3. Stepwise regression may be seen as mechanistic, potential non-linear terms were not tested for – however, findings are in line with earlier studies and post-hoc analyses using OLS with robust standard errors</li> </ol>
<b>Paper II: ATHENA WTA</b>	<ol style="list-style-type: none"> <li>1. Cost-effectiveness based on patient-level data (n=4628) with morbidity-mortality endpoints – relevant in payer and clinical decision-making</li> <li>2. Results stable across range of sensitivity analyses and assumptions, incl. post-hoc subgroup analyses for PALLAS</li> <li>3. Conservative estimates due to exclusion of e.g. indirect costs</li> </ol>	<ol style="list-style-type: none"> <li>1. Multinational trial covering different health care systems – however, heterogeneity test not significant</li> <li>2. Lack of utility data from trial population – addressed by adjusted results from Paper I</li> <li>3. Full-sample estimation did not adjust for right-censoring of cost data, leading to underestimation of overall costs (Raikou &amp; McGuire, 2004). As follow-up did not differ between treatment arms, the higher event rate in the placebo group would imply a bias against dronedarone. In similar analysis for AFFIRM, full-sample estimation did not change cost-effectiveness outcomes, but underestimated uncertainty of decision (Fenwick et al., 2008)</li> </ol>
<b>Paper III: Utilities in CHF</b>	<ol style="list-style-type: none"> <li>1. Large sample with national coverage (n=5334)</li> <li>2. New information on important patient subgroups, incl. different levels of LVEF</li> <li>3. Comprehensive statistical modelling approach, incl. results based on two different value sets (Sweden and UK)</li> </ol>	<ol style="list-style-type: none"> <li>1. Predominance of hospitalised, i.e. more severe, patients – adjustment for e.g. NYHA class in analyses</li> <li>2. Lack of comprehensive data on changes in clinical parameters and events after inclusion</li> <li>3. Generic HRQoL not as responsive to clinical change as disease-specific instruments – however, reflects clinically relevant differences between groups (e.g. NYHA class)</li> </ol>

<b>Paper IV: HRQoL and prognosis in CHF</b>	1. Long-term follow-up (8-12 years) of mortality and rehospitalisations	1. Small sample (n=177)
	2. Well characterised sample in terms of clinical variables, neurohormonal activation and echocardiographic measures	2. Patient sample treated with optimised drug therapy as of 1996-1999 – inherent to retrospective long-term studies
	3. Factor analysis allowed study of impact of specific HRQoL dimensions on outcomes	3. NHP not often used today – however, results reflective of other HRQoL instrument findings, incl. disease-specific measures

*Notes: AF=atrial fibrillation, CHF=chronic heart failure, CLAD=censored least absolute deviations, HRQoL=health-related quality of life, HTA=health technology assessment, LVEF=left ventricular ejection fraction, NHP=Nottingham Health Profile, NYHA=New York Heart Association, OLS=ordinary least squares*

### 6.1.1 Representativeness of study samples

An important aspect in the use of these study findings is the representativeness of the study samples for the patient populations of interest. In the Euro Heart Survey underlying Paper I, patients were recruited from cardiology clinics chosen to represent local medical practice (Nieuwlaat et al., 2005). A large share of patients was included in university centres (46%) and specialised centres (14%), indicating an overrepresentation of highly specialised clinics and those with an interest in AF research, which is also mirrored in the fact that 10-15% of patients were participating in clinical trials. Overall, 86% of patients had at least one stroke risk factor and 70% had AF symptoms at inclusion despite treatment. This could mean that our estimated utilities may have been lower than for an AF population including also less ill patients. However, patients filling in the EQ-5D at baseline were on average younger and healthier than those with incomplete information, which in turn may have led to an underestimation of the effect of different covariates. Moreover, when comparing to a wider AF population, patients seen in primary care are generally older and more frequently have permanent AF than those in a more specialised setting (Kirchhof et al., 2014), which might imply that patients from primary care who were not covered by the survey not necessarily would have had higher utilities.

When applying the utilities at follow-up from Paper I in the trial-based evaluation of ATHENA in Paper II, we adjusted for some of the known differences between the populations through available covariates, e.g. regarding age (mean age five years higher in ATHENA), gender (5% fewer men in ATHENA) and underlying heart disease (e.g. 21% with CHF history in ATHENA vs. 34% in Euro Heart Survey) (Hohnloser et al., 2009; Nieuwlaat et al., 2005). However, permanent AF was excluded in ATHENA, which may have led to an underestimation of baseline utility when applying the utility weights from Paper I. Furthermore, patients recruited into the trial are by nature usually not representative of patients seen in clinical practice. For example, the relatively high age limit in ATHENA (at least 70-75 years, depending on risk factors) means that in participating countries with lower average life expectancy, e.g. Russia, there would have been a clear overrepresentation of “healthy survivors” in the trial. This would probably not have occurred to the same extent in the observational Euro Heart Survey. Notwithstanding differences in study methods and patient characteristics, mortality outcomes in the two samples were relatively similar. In the



Euro Heart Survey, 1-year mortality was 5.3%, being highest in permanent AF (8.2%) and lowest in paroxysmal and persistent AF (3.5% and 3%, respectively) (Nieuwlaet et al., 2008). By comparison, in the high risk population with paroxysmal and persistent AF seen in ATHENA, 5-6% of patients died during the trial (mean duration 21 months) (Hohnloser et al., 2009).

The patient sample used for the analyses in Paper III was largely based on patients included in hospitals or in specialised outpatient care in Sweden. However, a recent Swedish registry analysis estimated that about 17% of HF patients are managed solely in primary care (Zarrinkoub et al., 2013). In reference to this population-based study, our sample is probably not representative of patients in the oldest age groups, women in general, or those managed only in primary care. The mix of comorbidities was also different in our sample. Finally, due to severity, NYHA class IV patients were also most likely underrepresented in our sample. These factors could have led to an overestimation of utilities for an overall CHF population in terms of age, gender and functional status, and possibly an underestimation due to the predominant inclusion from hospitals.

Patients in Paper IV were part of a randomised study evaluating a nurse-monitored management programme at a university hospital in Sweden during 1996-99. The patient sample focused on HFREF in NYHA classes II-IV, and received optimised medical therapy at the time of study. In comparison to current practice, patients had lower treatment rates with beta-blockers and renin angiotensin system blockers (Dahlstrom, Edner & Jonsson, 2012). The study is thus not suited for predicting mortality in a HFPEF population or those treated with contemporary drugs, where the relationship between HRQoL or utilities and long-term outcomes requires further study.

### **6.1.2 Meaningfulness of health-related quality of life measurements**

For applications in their work, clinicians and health care decision-makers need information on the clinical relevance of HRQoL results for differentiating between appropriate patient groups and detecting changes within groups or individuals over time. In this context, the responsiveness of HRQoL instruments should be taken into consideration.

Inherently, disease-specific instruments are more responsive to clinical changes than generic instruments; however, the latter may be more suited for the study context, provided that they also can capture important differences between groups or over time. In HFREF, the KCCQ has been shown to be more sensitive to clinical changes than the RAND12 (similar to the SF-12) physical and mental component scores and the EQ-5D, when using either the UK or US value sets (Eurich et al., 2006). A review of studies investigating the association between HRQoL and morbidity and mortality in CHF and coronary artery disease concluded that in CHF, physical health status measured with disease-specific instruments was better suited for outcome prediction than generic instruments due to fewer insignificant results (including SF-36, NHP, EQ-5D); however, in coronary artery disease, there was no difference between instrument types (Mommersteeg et al., 2009). In contrast, based on our analyses in Paper IV,

we found that the generic NHP instrument, and especially its physical domain, provided significant prognostic value both for mortality and rehospitalisations, even in our comparatively small sample. Interestingly, the impact of HRQoL on the hazard for mortality in our study was similar to the one seen in a larger longitudinal study using the MLWHFQ (Lupon et al., 2013), which may indicate that the measured impact is similar for the two generic and disease-specific instruments. In a cross-sectional study of HF outpatients, a relatively high correlation was found between MLWHFQ summary and EQ-5D index scores (de Rivas et al., 2008), which suggests that the EQ-5D may be a valid and acceptable method for measuring HRQoL in these patients, also owing to its simple and fast administration. Our results in Paper III support and extend this finding, as we saw a clear distinction between e.g. NYHA classes when using the EQ-5D in mainly hospitalised patients.

To our knowledge, there are no studies evaluating generic vs. disease-specific instruments in AF that provide a similar context for our findings in Paper I. Instead, previously established minimally important differences for the EQ-5D may be used as a reference point. For the UK value set, the mean minimally important difference has been estimated at 0.08 using instrument-defined health transitions (Luo, Johnson & Coons, 2010). In this light, our results in Paper I would mean that clinically relevant differences were found especially for symptom burden and major clinical comorbidities at baseline and events during follow-up (e.g. stroke, myocardial infarction, CHF).

When comparing different generic HRQoL instruments in patient groups similar to those in our studies, previous research in rehabilitation patients with relatively mild conditions, including cardiovascular disorders, found that the EQ-5D was less responsive to change than the SF-6D, despite a wider spread of health state values. This was deemed to be a likely reflection of the smaller amount of possible health states when using the EQ-5D (Moock & Kohlmann, 2008). We could not identify any studies testing the responsiveness of the NHP in cardiovascular patients. However, in elderly women with femoral neck fractures, a comparison with the EQ-5D showed that both instruments had good internal and external responsiveness, while the EQ-5D had the advantage of already providing sufficient information when used on its own for discriminating between patients according to the external clinical criterion of fracture displacement (Tidermark & Bergstrom, 2007).

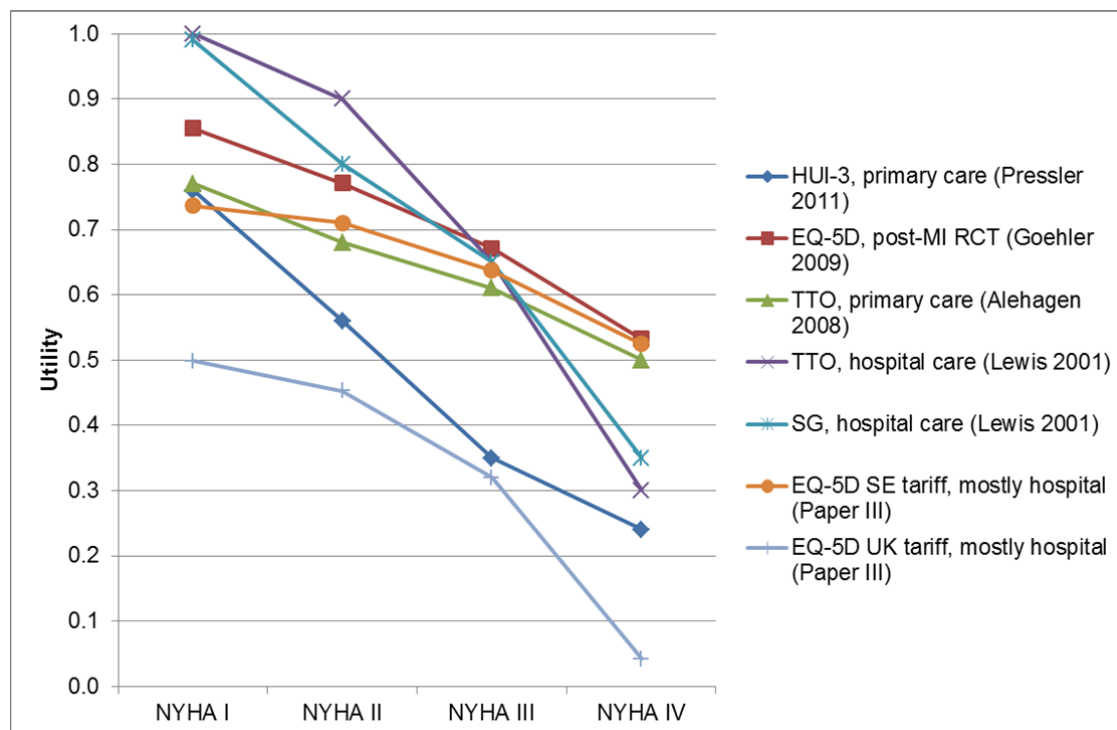
In summary, we found that the generic EQ-5D and NHP instruments were able to distinguish between different groups of patients, also over time. In AF, the EQ-5D was able to detect differences between different symptoms and comorbidities that can be deemed as clinically relevant, both at baseline and after one year (Paper I). In CHF, we also found that the EQ-5D was able to distinguish between clinically relevant patient groups (Paper III). Moreover, the NHP instrument allowed stratification of patients in terms of prognosis, both when measured at baseline hospitalisation (Paper IV) and at 1-year follow-up in outpatient care (Section 5.4.1). Depending on the type and amount of HRQoL information required, these generic instruments can thus provide informative results for use both in clinical practice and in economic evaluations.

### 6.1.3 Considerations when using utilities from different studies

As outlined in Section 2.2.1.2, different HRQoL instruments and valuation techniques lead to different utility estimates. This also needs to be borne in mind when using the results from Papers I and III. As an illustration, Figure 21 summarises utility values by NYHA class in CHF from previous studies (described in Section 2.3.3) and Paper III. Several observations can be made based on this:

- In all studies, there is a clear inverse association between increasing NYHA class and utilities;
- Utilities obtained with the same method in different samples, e.g. time trade-off (Alehagen et al., 2008; Lewis et al., 2001) or EQ-5D (Gohler et al., 2009; Paper III), do not necessarily have the same levels or follow the same slope;
- Conversely, utilities derived in different settings and based on different value sets or valuation techniques may still yield similar estimates (e.g. Alehagen et al., 2008; Gohler et al., 2009 and Swedish tariff in Paper III);
- Different value sets for the EQ-5D can have a large impact on utilities (Paper III);
- Methods allowing for negative utilities (HUI-3, UK tariff for EQ-5D) tend to lead to lower utilities across health states.

**Figure 21: Comparison of utilities by NYHA class in CHF for different types of elicitation methods and samples**



Notes: Paper III – utilities derived for men, 70-79 years, other variables at reference level

These observations represent the first comparison of existing utility values in CHF. They add to other research around the impact of measurement and analysis options on health state

utilities. The impact of value sets derived in different countries and through different techniques has been demonstrated both on a general level (Knies et al., 2009) and for specific diseases (e.g. in rheumatoid arthritis, see Karlsson et al., 2011). Even when using value sets from the same country, the choice of instrument and scoring algorithm has been shown to impact on the predicted quality-adjusted survival, both in terms of absolute levels and in terms of incremental changes (Sorensen et al., 2012). The implications of selecting a specific value set can thus be profound, as the slope and level of the utility curves across health states has a direct impact on QALYs associated with different treatment strategies. For example, the use of a different value set with a smaller slope in Paper I would most likely have led to lower incremental QALYs in Paper II. Instead, when comparing dronedarone to other interventions in the same indication, the effect of treatment on survival would have received a proportionally larger weight.

## **6.2 THE ROLE OF HEALTH-RELATED QUALITY OF LIFE AND UTILITIES IN PUBLIC HEALTH**

### **6.2.1 Health policy applications: use by payers**

#### *6.2.1.1 Established use in payer decisions – example of dronedarone*

In many countries, HRQoL and utilities are part of the evidence considered by decision-makers in reimbursement authorities. On one end of the spectrum, they may be seen as part of the clinical evidence, in that they are potential outcome measures for clinical effectiveness. On the other end, utilities can be required to obtain QALYs as a summary measure in an economic evaluation. In this thesis, Papers I, II and III provide evidence that would be used in the latter type of decision-making framework. Papers I and III can be used as sources for economic evaluations, while Paper II illustrates a direct application of the findings from Paper I that was part of the health economic evidence for dronedarone. The trial-based analysis was complemented by modelling studies (e.g. Akerborg et al., 2012; Nilsson et al., 2013), which also used findings from Paper I. During the analyses for Paper II, new safety results became available for dronedarone as part of the PALLAS trial, which led to a restriction of its indication. To reflect this, a post-hoc subgroup analysis was performed on patients in ATHENA, which provided important sensitivity results.

In May 2010, the Canadian Expert Drug Advisory Committee (CEDAC) finally recommended that dronedarone should not be listed. Its reasons were the uncertainty of the therapeutic benefit of dronedarone compared to amiodarone, together with a higher price, lack of positive rhythm outcomes and lack of HRQoL data from the clinical trial programme (Canadian Agency for Drugs and Technologies in Health, 2010). This decision was in contrast to the treatment recommendations by the Canadian Cardiovascular Society and European Society of Cardiology, even after the restrictions in the use of dronedarone.

NICE in the UK has closely integrated QALYs into its decision-making. In its preliminary guidance, NICE rejected dronedarone due to lower effectiveness and higher costs compared to existing AF treatments (McKenna et al., 2012). However, the subsequent consultation

process with clinical experts and patient representatives showed strong disagreement with the initial decision, and highlighted the need for clinical alternatives to available antiarrhythmic treatment. This led to a final recommendation by NICE for dronedarone to be used as a second-line option for patients with non-permanent AF and at least one cardiovascular risk factor.

In the NICE appraisal, the Evidence Review Group commented on the utility data used in the UK model. It voiced concerns about the application of utility weights from a single source (Euro Heart Survey), and specifically the process of arriving at this source, since there was no reporting of a systematic review and the values were not placed into a context of other studies or UK population norms. When comparing to population norms, the Evidence Review Group saw a risk of potential overestimation of the treatment effect of dronedarone. Therefore, it tested to estimate utility decrements of specific health states instead and applied these to UK population values. This sensitivity analysis based on a post-hoc adjustment turned out to have a limited impact on the ICER (McKenna et al., Centre for Reviews and Dissemination and Centre for Health Economics, 2009).

In the reimbursement decision by the Swedish TLV, the agency followed the indication for dronedarone, limiting it to patients with non-permanent AF and at least one cardiovascular risk factor (previous stroke or transient ischaemic attack, hypertension, diabetes, age above 75 years). In addition, TLV requested follow-up with real-world evidence on concomitant treatment, stroke risk assessment and clinical outcomes (Tandvårds- och Läkemedelsförmånsverket, 2011). In relation to the Canadian example, it is interesting to note that treatment guidelines in Sweden take health economic findings into account, where the cost per QALY is one of several factors weighing into an overall priority grading for different treatments in a given indication (Grip et al., 2011).

The Institute for Quality and Efficiency in Health Care (IQWiG) in Germany is an example of an authority that does not primarily use utilities or QALYs in its decision-making. Instead of effectiveness, IQWiG uses the term “benefit”, which is assessed in relationship to costs (Institute for Quality and Efficiency in Health Care, 2009). Benefit may be defined in a more narrow sense (medical benefit) or a wider sense (including “value” aspects). IQWiG does not recommend any specific instrument or procedure for measuring benefits. Since economic evaluations are not used for allocating resources across different therapeutic areas in Germany, it is possible to use different benefit measures in different therapeutic areas. Benefit may be quantified using clinical (including HRQoL), responder or aggregated measures. QALYs are mentioned as one potential example of aggregated measures, which may be used within an indication if deemed suitable. Rather than WTP thresholds for QALYs, IQWiG uses an “efficiency frontier” by therapeutic area in its decision-making. Since IQWiG has not published any evaluation of dronedarone, it is not possible to state what specific evidence regarding benefits would have been considered in this case. In general, HRQoL measures or utilities could be used for an indication-specific assessment; at the same

time, it would also be possible to use non-preference based HRQoL measures to demonstrate benefit.

#### *6.2.1.2 Newer applications in quality of care initiatives*

An interesting development is the use of QALYs as an aggregate measure to assess quality of care. Traditionally, this has been performed on the basis of selected process indicators, e.g. number of tests performed per patient and year, or clinical indicators, such as mortality. Some examples of the wider application of QALYs have been found in the US. In the area of diabetes risk factor control, QALYs were used to assess therapeutic value across different medical centres and were found to be a more useful and informative measure than levels of control for individual risk factors (Schmittiel et al., 2008). In a study on the quality of acute hospital care in coronary artery bypass grafting, utility weights were integrated with more traditional performance measures (e.g. mortality, complications) to derive a composite QALY measure that was able to take different surgical complications and long-term outcomes into account (Timbie et al., 2009).

In a nationwide context, HRQoL has since 2009 been included as part of standard data collection for pre- and post-surgery (knee, hip, hernia, varicose veins) in the UK. The EQ-5D and disease-specific measures have been recommended as patient-reported outcomes measures. This initiative is also being extended to other indications (Black, 2013). The information can be used to aid in clinical management of individual patients, assessment and comparison of health care providers and for evaluating treatment practices within a given indication.

In Sweden, patient-reported outcomes measures are collected as part of several of the numerous quality registers (such as the Swedish Heart Failure Registry used in Paper III). In 2014, 38 quality registers were using the EQ-5D as part of their regular data collection (not counting pilot studies), while seven were using the SF-36 and 16 were using disease-specific measures (PROM center, 2014). In a recent report for the National Board of Health and Welfare (Nilsson, 2014), no recommendations were given yet as to which measures should be used, because it was deemed that further development work and critical evaluation was needed, e.g. on the value of national comparisons, and different methodological options regarding instruments and analysis. As a case in point, the Swedish Heart Failure Registry is collaborating with the national Patient Reported Outcome Measures (PROM) centre in Linköping to further develop the methods around the use of patient-reported outcomes in this registry.

In light of these developments, many methodological questions become relevant, e.g. regarding suitable instruments, valuation methods and analysis of outcomes. Since the wide range of available approaches can lead to markedly different results, development of a “reference case” as suggested for the QALY (Drummond et al., 2009) would seem to be crucial, even more so because these are challenging concepts to communicate to wider, non-specialist audiences. As it is likely that such a reference case would have to remain relatively

general, normative choices will probably still be required at a decision-maker level (in Sweden e.g. TLV, National Board of Health and Welfare, regional drug committees).

### **6.2.2 Prognostic markers in clinical practice**

In clinical practice, potential applications of QALYs could be to measure patient health status, compare individuals against reference populations and measure change over time in individuals and patient groups (Kind et al., 2009). This is certainly also true for HRQoL.

Prognostic factors may also be used in economic evaluations to stratify treatment effectiveness. For example, in the economic models for dronedarone, the CHADS<sub>2</sub> score was used to stratify treatment effectiveness on all-cause mortality by different stroke risk levels (Akerborg et al., 2012; McKenna et al., 2012; Nilsson et al., 2013).

While the relevance of HRQoL for prognosis has been quite extensively studied in CHF, this has not been the case in AF. Recently published analyses from a QoL sub-study of AFFIRM (Schron, Friedmann & Thomas, 2014) showed that HRQoL measured with the generic SF-36 instrument, and in particular its physical component score, was a significant predictor of both 1-year mortality and hospitalisation in AF patients. In addition, the mental component score was a separate predictor for hospitalisations. Thus, the prognostic value of HRQoL is worth further research in AF, and probably also in other chronic indications.

In light of the findings of Paper IV and previous research, it could be worth to explore possibilities for including HRQoL in risk prediction models to a larger extent than has been done to date. For wider use in clinical practice, the following aspects should be considered:

- The HRQoL scale should be able to capture clinically important changes, as discussed in Section 6.1.2;
- The prediction parameters should be quite easily available as part of standard care, which in some countries may see an increasing uptake of HRQoL information through e.g. quality of care initiatives as described in the previous section;
- A comprehensive set of significant predictors, including clinical parameters, e.g. physician assessments or diagnostic tests, may improve the precision of the predicted outcome;
- Risk prediction models may be more suited for “pure” clinical endpoints such as mortality, as e.g. readmissions are more likely to depend on local health care organisation, funding and treatment practices.

For implementation in clinical practice, standard procedures could be introduced to measure HRQoL (e.g. using the EQ-5D) for risk stratification and subsequent follow-up of high-risk patients.

### **6.2.3 Possible use with patients**

The studies in this thesis have not dealt with direct use of HRQoL or utilities with individual patients. Since it is patients’ HRQoL that is investigated, the information can form a natural

component when involving patients in evaluation of and decisions regarding their health. Shared decision-making aims to promote collaboration between patients and clinicians by taking into account best available evidence together with individual patient characteristics and preferences (Politi et al., 2013). This is usually focused on clinical comparative effectiveness (benefits and risks) of different treatment alternatives. An interesting example of possible extensions of this concept to a decision-support system involving utilities is a mobile tool that has been tested in a small group of AF patients (Quaglini et al., 2013). The tool uses techniques from economic evaluations as part of its decision-making framework, including decision trees, Markov models, and patient preference for different health states using standard gamble or time trade-off techniques. In this context, one needs to consider potential challenges when communicating about concepts such as QALYs, probabilities, uncertainty of outcomes and the difference between population means and individual disease trajectories.

### **6.3 SUGGESTIONS FOR FUTURE RESEARCH**

Based on the studies covered in this thesis and their place in the wider public health context, there are several areas that would warrant further research:

- Compare different generic and disease-specific HRQoL instruments in the same patient sample to better understand their responsiveness and ability to capture relevant domains – especially in AF, there is a lack of such studies to date, and it would be interesting to evaluate some of the recently developed AF-specific instruments in this respect;
- Investigate the value of generic instruments, e.g. the EQ-5D, in detecting change over time and providing prognostic information in broad patient samples – this would be of interest both in CHF and in AF and could include outcomes such as mortality and disease progression;
- Investigate the relationship between HRQoL/utilities and other covariates in specific patient groups for more precise estimates in economic evaluations – in CHF, the link between LVEF and HRQoL/utilities is of particular interest; other parameters could include disease type (e.g. in AF), symptom levels (also in AF), comorbidities, care setting (primary vs. specialist care);
- Test the development of a risk scoring tool that includes HRQoL together with key clinical information for patient groups of interest – this would require external validation and could then be used to classify patients according to risk in clinical practice and in the evaluation of new therapies;
- Understand how HRQoL and utilities are used in the health care system and clinical decision-making – ideally, the goal would be to generate a common understanding of the issues in terms of methods, implementation and communication and how these could be approached going forward, e.g. in the context of quality of care initiatives.



In general, when deciding on the most suitable study design to address these questions, different alternatives should be evaluated in terms of both the type and representativeness of data that can be collected and the associated costs and time requirements. Potential options include e.g. cross-sectional or longitudinal data collection, as well as use of data from national or local registries.



## 7 CONCLUSIONS

In summary, the studies in this thesis support the use of HRQoL and utilities as a value-added part of clinical and economic decision-making, due to their relationship with both clinical and economic outcomes. Specifically, the following conclusions can be drawn:

- In AF patients treated in specialist care, increasing age, female gender, domestic status outside the own home, existing comorbidities, and symptoms of AF, CHF or angina were associated with reduced utility at baseline, while regular physical activity had a positive effect. At 1-year follow-up, significant determinants included AF symptoms and major adverse events, including stroke, myocardial infarction and CHF (Paper I);
- The information on determinants of utility can be used in different types of economic evaluations, as shown in the cost-effectiveness evaluations of dronedarone (including Paper II). The within-trial analysis indicated that dronedarone when used as in ATHENA is cost-effective within generally accepted thresholds (base case ICER CAD\$7560 per QALY), and that this also would hold in the light of subsequent label restrictions (Paper II);
- In CHF patients mainly seen in hospital, utility at baseline was negatively affected by female gender, increasing age, increasing NYHA class, preserved LVEF, lung disease, diabetes, and use of nitrates, antiplatelets or diuretics. Higher systolic blood pressure and haemoglobin levels and use of angiotensin converting enzyme inhibitors/angiotensin receptor blockers or beta-blockers were associated with increased utility. A significant interaction between age category and NYHA class shows that patients in the youngest age group are more severely affected by worsening NYHA status than older patients. In our data set, the OLS model performed slightly better than the two-part model on a population level and for capturing utility ranges (Paper III);
- In patients hospitalised with systolic CHF, physical mobility as measured with a generic HRQoL instrument was a significant independent predictor for all-cause and cardiovascular rehospitalisation and mortality, with every 1% worsening resulting in a 1-2% increase in the hazard ratio of being hospitalised or dying. Emotional reactions were an additional independent predictor for all-cause hospitalisations, with a similar impact as physical mobility (Paper IV). Additional analyses suggest that the impact of HRQoL, and specifically physical mobility, on cardiovascular mortality may be similar irrespective of timing and setting of assessment.



## 8 ACKNOWLEDGEMENTS

I would like to express my deep gratitude to the following people, who have all contributed to the generation of this work in various ways:

- Peter Lindgren, my main supervisor, for agreeing to be my supervisor and always trusting in my capabilities, for your intelligent and creative input, and for sharing your excellent knowledge of health economics;
- Thomas Kahan, my supervisor, for your positive and inspiring outlook, constructive input throughout, strong dedication to my research development, and for generously sharing your extensive clinical expertise;
- Carolin Miltenburger, my supervisor, for friendly collaboration over the years, providing a complementary perspective on the application of health economics and outcomes research in industry;
- My co-authors, Robby Nieuwlaat, Harry Crijns, Stuart Connolly, Märit Mejhert, Hans Persson, Magnus Edner and Ulf Dahlström, for friendly and knowledgeable collaboration, providing highly valued insights and guidance on clinical matters;
- Journal club members at the Institute of Environmental Medicine, Tony Bohman, Jessica Magnusson, Kari Paanalahti, Lina Palmlöf, David Pettersson and Dong Wang, for a welcoming atmosphere and helpful discussions around epidemiology, statistical methods and life as a PhD student;
- Linus Jönsson for hiring me back in the Stockholm Health Economics days, for sharing your vast health economic expertise and providing an inspirational work environment;
- Colleagues past and present (at Stockholm Health Economics/European Health Economics/i3 Innovus/OptumInsight) for many enriching discussions about health economics and a fun and stimulating work atmosphere;
- Jonas Lundkvist for helpful comments on the thesis summary;
- Jenny Lanyon for highly valuable mentorship, inspiration and support;
- My friends from many places and times for your friendship and encouragement;
- My family-in-law, Anne-Marie and Fernand, Marie-Anne and Laurent, Pascale and Pierre, Catherine and Emmanuel, for your kindness and generosity;
- My parents, Catarina and Jan, for your love, encouragement and for passing on the pleasure in learning;
- Xavier, Esther and Aurel, for your love, support and the joy of being with you.

Financial support from Sanofi-Aventis and Novartis for several of the studies is gratefully acknowledged.



## 9 REFERENCES

- Akerborg O, Nilsson J, Bascle S, Lindgren P & Reynolds M (2012). Cost-effectiveness of dronedarone in atrial fibrillation: results for Canada, Italy, Sweden, and Switzerland. *Clin Ther*; 34: 1788-1802.
- Akintade BF, Chapa D, Friedmann E & Thomas SA (2013). The influence of depression and anxiety symptoms on health-related quality of life in patients with atrial fibrillation and atrial flutter. *J Cardiovasc Nurs*.
- Alam M, Bandeali SJ, Shahzad SA & Lakkis N (2012). Real-life global survey evaluating patients with atrial fibrillation (REALISE-AF): results of an international observational registry. *Expert Rev Cardiovasc Ther*; 10: 283-291.
- Alba AC, Agoritsas T, Jankowski M, Courvoisier D, Walter SD, Guyatt GH et al. (2013). Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. *Circ Heart Fail*; 6: 881-889.
- Alehagen U, Rahmqvist M, Paulsson T & Levin LA (2008). Quality-adjusted life year weights among elderly patients with heart failure. *Eur J Heart Fail*; 10: 1033-1039.
- Arribas F, Ormaetxe JM, Peinado R, Perulero N, Ramirez P & Badia X (2010). Validation of the AF-QoL, a disease-specific quality of life questionnaire for patients with atrial fibrillation. *Europace*; 12: 364-370.
- Ball J, Carrington MJ, McMurray JJ & Stewart S (2013). Atrial fibrillation: profile and burden of an evolving epidemic in the 21st century. *Int J Cardiol*; 167: 1807-1824.
- Barasa A, Schaufelberger M, Lappas G, Swedberg K, Dellborg M & Rosengren A (2014). Heart failure in young adults: 20-year trends in hospitalization, aetiology, and case fatality in Sweden. *Eur Heart J*; 35: 25-32.
- Berg J, Lindgren P, Nieuwlaat R, Bouin O & Crijns H (2010). Factors determining utility measured with the EQ-5D in patients with atrial fibrillation. *Qual Life Res*; 19: 381-390.
- Berg J, Sauriol L, Connolly S & Lindgren P (2013). Cost-effectiveness of dronedarone in patients with atrial fibrillation in the ATHENA trial. *Can J Cardiol*; 29: 1249-1255.
- Berg J, Lindgren P, Kahan T, Schill O, Persson H, Edner M et al. (2014). Health-related quality of life and long-term morbidity and mortality in patients hospitalised with systolic heart failure. *JRSM Cardiovasc Dis*; 3: 2048004014548735.
- Berry C & McMurray J (1999). A review of quality-of-life evaluations in patients with congestive heart failure. *Pharmacoeconomics*; 16: 247-271.
- Black N (2013). Patient reported outcome measures could help transform healthcare. *BMJ*; 346: f167.
- Bleumink GS, Knetsch AM, Sturkenboom MC, Straus SM, Hofman A, Deckers JW et al. (2004). Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure. The Rotterdam Study. *Eur Heart J*; 25: 1614-1619.
- Brazier J, Roberts J & Deverill M (2002). The estimation of a preference-based measure of health from the SF-36. *J Health Econ*; 21: 271-292.
- Brazier J, Roberts J, Tsuchiya A & Busschbach J (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*; 13: 873-884.
- Brooks R (1996). EuroQol: the current state of play. *Health Policy*; 37: 53-72.
- Brouwer WB, Culyer AJ, van Exel NJ & Rutten FF (2008). Welfarism vs. extra-welfarism. *J Health Econ*; 27: 325-338.

- Burström K, Johannesson M & Diderichsen F (2001). Health-related quality of life by disease and socio-economic group in the general population in Sweden. *Health Policy*; 55: 51-69.
- Burström K, Sun S, Gerdtham UG, Henriksson M, Johannesson M, Levin LA et al. (2014). Swedish experience-based value sets for EQ-5D health states. *Qual Life Res*; 23: 431-442.
- Calvert MJ, Freemantle N & Cleland JG (2005). The impact of chronic heart failure on health-related quality of life data acquired in the baseline phase of the CARE-HF study. *Eur J Heart Fail*; 7: 243-251.
- Camm AJ, Kirchhof P, Lip GY, Schotten U, Savelieva I, Ernst S et al. (2010). Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur Heart J*; 31: 2369-2429.
- Camm AJ, Lip GY, De CR, Savelieva I, Atar D, Hohnloser SH et al. (2012). 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. *Eur Heart J*; 33: 2719-2747.
- Canadian Agency for Drugs and Technologies in Health (2006). *Guidelines for the economic evaluation of health technologies: Canada (3rd edition)*. Ottawa.
- Canadian Agency for Drugs and Technologies in Health (2010). *CEDAC Final recommendation: dronedarone hydrochloride*.
- Caro JJ, Briggs AH, Siebert U & Kuntz KM (2012). Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health*; 15: 796-803.
- Chamberlain AM, McNallan SM, Dunlay SM, Spertus JA, Redfield MM, Moser DK et al. (2013). Physical health status measures predict all-cause mortality in patients with heart failure. *Circ Heart Fail*; 6: 669-675.
- Chan PS, Vijan S, Morady F & Oral H (2006). Cost-effectiveness of radiofrequency catheter ablation for atrial fibrillation. *J Am Coll Cardiol*; 47: 2513-2520.
- Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ et al. (2014). Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation*; 129: 837-847.
- Coons SJ, Rao S, Keininger DL & Hays RD (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics*; 17: 13-35.
- Cowie MR, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Suresh V et al. (1999). Incidence and aetiology of heart failure; a population-based study. *Eur Heart J*; 20: 421-428.
- Coyne K, Margolis MK, Grandy S & Zimetbaum P (2005). The state of patient-reported outcomes in atrial fibrillation : a review of current measures. *Pharmacoeconomics*; 23: 687-708.
- Dahlström U, Edner M & Jonsson A (2012). Årsrapport RiksSvikt: 2011 års resultat.
- de Leon CF, Grady KL, Eaton C, Rucker-Whitaker C, Janssen I, Calvin J et al. (2009). Quality of life in a diverse population of patients with heart failure: Baseline findings from the Heart Failure Adherence and Retention Trial (HART). *J Cardiopulm Rehabil Prev*; 29: 171-178.



de Rivas B, Permanyer-Miralda G, Brotons C, Aznar J & Sobreviela E (2008). Health-related quality of life in unselected outpatients with heart failure across Spain in two different health care levels. Magnitude and determinants of impairment: the INCA study. *Qual Life Res*; 17: 1229-1238.

Devlin NJ & Krabbe PF (2013). The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ*; 14 Suppl 1: S1-S3.

Dolan P (1997). Modeling valuations for EuroQol health states. *Med Care*; 35: 1095-1108.

Dolders MG, Zeegers MP, Groot W & Ament A (2006). A meta-analysis demonstrates no significant differences between patient and population preferences. *J Clin Epidemiol*; 59: 653-664.

Dorian P, Paquette M, Newman D, Green M, Connolly SJ, Talajic M et al. (2002). Quality of life improves with treatment in the Canadian Trial of Atrial Fibrillation. *Am Heart J*; 143.

Drummond M, Sculpher M, Torrance G, O'Brien BJ & Stoddart G (2005). *Methods for economic evaluation of health care programmes*. (Third ed.) Oxford: Oxford University Press.

Drummond M, Brixner D, Gold M, Kind P, McGuire A & Nord E (2009). Toward a consensus on the QALY. *Value Health*; 12 Suppl 1: S31-S35.

Dyer MT, Goldsmith KA, Sharples LS & Buxton MJ (2010). A review of health utilities using the EQ-5D in studies of cardiovascular disease. *Health Qual Life Outcomes*; 8: 13.

Eichler HG, Kong SX, Gerth WC, Mavros P & Jonsson B (2004). Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? *Value Health*; 7: 518-528.

Eurich DT, Johnson JA, Reid KJ & Spertus JA (2006). Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health Qual Life Outcomes*; 4: 89.

Faller H, Stork S, Schowalter M, Steinbuchel T, Wollner V, Ertl G et al. (2007). Is health-related quality of life an independent predictor of survival in patients with chronic heart failure? *J Psychosom Res*; 63: 533-538.

Faller H, Stork S, Schuler M, Schowalter M, Steinbuchel T, Ertl G et al. (2009). Depression and disease severity as predictors of health-related quality of life in patients with chronic heart failure--a structural equation modeling approach. *J Card Fail*; 15: 286-292.

Fayers PM & Machin D (2007). *Quality of life. The assessment, analysis and interpretation of patient-reported outcomes*. (Second ed.) Chichester: John Wiley & Sons Ltd.

Fenwick E, O'Brien BJ & Briggs A (2004). Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. *Health Econ*; 13: 405-415.

Fenwick E, Marshall DA, Blackhouse G, Vidaillet H, Slee A, Shemanski L et al. (2008). Assessing the impact of censoring of costs and effects on health-care decision-making: an example using the Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) study. *Value Health*; 11: 365-375.

Franzen K, Saveman BI & Blomqvist K (2007). Predictors for health related quality of life in persons 65 years or older with chronic heart failure. *Eur J Cardiovasc Nurs*; 6: 112-120.

Furlong WJ, Feeny DH, Torrance GW & Barr RD (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med*; 33: 375-384.

Gafni A & Birch S (2006). Incremental cost-effectiveness ratios (ICERs): the silence of the lambda. *Soc Sci Med*; 62: 2091-2100.

Garin O, Ferrer M, Pont A, Rue M, Kotzeva A, Wiklund I et al. (2009). Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. *Qual Life Res*; 18: 71-85.

Garrison LP, Jr. (2009). Editorial: On the benefits of modeling using QALYs for societal resource allocation: the model is the message. *Value Health*; 12 Suppl 1: S36-S37.

Gerdtham UG & Zethraeus N (2003). Predicting survival in cost-effectiveness analyses based on clinical trials. *Int J Technol Assess Health Care*; 19: 507-512.

Gohler A, Geisler BP, Manne JM, Kosiborod M, Zhang Z, Weintraub WS et al. (2009). Utility estimates for decision-analytic modeling in chronic heart failure--health states based on New York Heart Association classes and number of rehospitalizations. *Value Health*; 12: 185-187.

Gott M, Barnes S, Parker C, Payne S, Seamark D, Gariballa S et al. (2006). Predictors of the quality of life of older people with heart failure recruited from primary care. *Age Ageing*; 35: 172-177.

Green C, Brazier J & Deverill M (2000). Valuing health-related quality of life. A review of health state valuation techniques. *Pharmacoeconomics*; 17: 151-165.

Grip L, Lindahl B, Levin LA, Karvinge C, Eklund K & Wallentin L (2011). From European to national guidelines on heart disease. *Scand Cardiovasc J*; 45: 3-13.

Groenveld HF, Crijns HJ, Van den Berg MP, Van Sonderen E, Alings AM, Tijssen JG et al. (2011). The effect of rate control on quality of life in patients with permanent atrial fibrillation: data from the RACE II (Rate Control Efficacy in Permanent Atrial Fibrillation II) study. *J Am Coll Cardiol*; 58: 1795-1803.

Gutzwiler FS, Pfeil AM, Comin-Colet J, Ponikowski P, Filippatos G, Mori C et al. (2013). Determinants of quality of life of patients with heart failure and iron deficiency treated with ferric carboxymaltose: FAIR-HF sub-analysis. *Int J Cardiol*; 168: 3878-3883.

Hagens VE, Ranchor AV, Van Sonderen E, Bosker HA, Kamp O, Tijssen JG et al. (2004a). Effect of rate or rhythm control on quality of life in persistent atrial fibrillation. Results from the Rate Control Versus Electrical Cardioversion (RACE) Study. *J Am Coll Cardiol*; 43.

Hagens VE, Vermeulen KM, TenVergert EM, van Veldhuisen DJ, Bosker HA, Kamp O et al. (2004b). Rate control is more cost-effective than rhythm control for patients with persistent atrial fibrillation--results from the RAtE Control versus Electrical cardioversion (RACE) study. *Eur Heart J*; 25: 1542-1549.

Hatoum HT, Brazier JE & Akhras KS (2004). Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting. *Value Health*; 7: 602-609.

Heeringa J, van der Kuip DA, Hofman A, Kors JA, van HG, Stricker BH et al. (2006). Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *Eur Heart J*; 27: 949-953.

Hoekstra T, Jaarsma T, van Veldhuisen DJ, Hillege HL, Sanderman R & Lesman-Leegte I (2013). Quality of life and survival in patients with heart failure. *Eur J Heart Fail*; 15: 94-102.

Hohnloser SH, Crijns HJ, van EM, Gaudin C, Page RL, Torp-Pedersen C et al. (2009). Effect of dronedarone on cardiovascular events in atrial fibrillation. *N Engl J Med*; 360: 668-678.

Holland R, Rechel B, Stepien K, Harvey I & Brooksby I (2010). Patients' self-assessed functional status in heart failure by New York Heart Association class: a prognostic predictor of hospitalizations, quality of life and death. *J Card Fail*; 16: 150-156.

Huang IC, Frangakis C, Atkinson MJ, Willke RJ, Leite WL, Vogel WB et al. (2008). Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. *Health Serv Res*; 43: 327-339.

Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J & Papp E (1980). A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health*; 34: 281-286.

Institute for Quality and Efficiency in Health Care (2009). *General methods for the assessment of the relation of benefits to costs*.

Iqbal J, Francis L, Reid J, Murray S & Denvir M (2010). Quality of life in patients with chronic heart failure and their carers: a 3-year follow-up study assessing hospitalization and mortality. *Eur J Heart Fail*; 12: 1002-1008.

Jhund PS, Macintyre K, Simpson CR, Lewsey JD, Stewart S, Redpath A et al. (2009). Long-term trends in first hospitalization for heart failure and subsequent survival between 1986 and 2003: a population study of 5.1 million people. *Circulation*; 119: 515-523.

Johansson P, Brostrom A, Dahlstrom U & Alehagen U (2008). Global perceived health and ten-year cardiovascular mortality in elderly primary care patients with possible heart failure. *Eur J Heart Fail*; 10: 1040-1047.

Kannel WB, Wolf PA, Benjamin EJ & Levy D (1998). Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates. *Am J Cardiol*; 82: 2N-9N.

Karlsson JA, Nilsson JA, Neovius M, Kristensen LE, Gulfe A, Saxne T et al. (2011). National EQ-5D tariffs and quality-adjusted life-year estimation: comparison of UK, US and Danish utilities in south Swedish rheumatoid arthritis patients. *Ann Rheum Dis*; 70: 2163-2166.

Kind P, Lafata JE, Matuszewski K & Raisch D (2009). The use of QALYs in clinical and patient decision-making: issues and prospects. *Value Health*; 12 Suppl 1: S27-S30.

Kirchhof P, Schmalowsky J, Pittrow D, Rosin L, Kirch W, Wegscheider K et al. (2014). Management of patients with atrial fibrillation by primary-care physicians in Germany: 1-year results of the ATRIUM registry. *Clin Cardiol*; 37: 277-284.

Kirkwood B & Sterne J (2005). *Medical Statistics*. (Second ed.) Blackwell Science.

Knies S, Evers SM, Candel MJ, Severens JL & Ament AJ (2009). Utilities of the EQ-5D: transferable or not? *Pharmacoeconomics*; 27: 767-779.

Kontodimopoulos N, Argiriou M, Theakos N & Niakas D (2011). The impact of disease severity on EQ-5D and SF-6D utility discrepancies in chronic heart failure. *Eur J Health Econ*; 12: 383-391.

Leal J, Luengo-Fernandez R & Gray A (2012). Economic costs. In Nichols M, Townsend N, Scarborough P, Rayner M & et al (Eds.), *European Cardiovascular Disease Statistics 2012*. European Society of Cardiology.

- Lee DT, Yu DS, Woo J & Thompson DR (2005). Health-related quality of life in patients with congestive heart failure. *Eur J Heart Fail*; 7: 419-422.
- Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB et al. (2006). The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*; 113: 1424-1433.
- Lewis EF, Johnson PA, Johnson W, Collins C, Griffin L & Stevenson LW (2001). Preferences for quality of life or survival expressed by patients with heart failure. *J Heart Lung Transplant*; 20: 1016-1024.
- Lip GY, Kakar P & Watson T (2007). Atrial fibrillation--the growing epidemic. *Heart*; 93: 542-543.
- Lloyd-Jones DM, Larson MG, Leip EP, Beiser A, D'Agostino RB, Kannel WB et al. (2002). Lifetime risk for developing congestive heart failure: the Framingham Heart Study. *Circulation*; 106: 3068-3072.
- Lloyd-Jones DM, Wang TJ, Leip EP, Larson MG, Levy D, Vasan RS et al. (2004). Lifetime risk for development of atrial fibrillation: the Framingham Heart Study. *Circulation*; 110: 1042-1046.
- Luo N, Johnson J & Coons SJ (2010). Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Med Care*; 48: 365-371.
- Lupon J, Gastelurrutia P, de AM, Gonzalez B, Cano L, Cabanes R et al. (2013). Quality of life monitoring in ambulatory heart failure patients: temporal changes and prognostic value. *Eur J Heart Fail*; 15: 103-109.
- Maggioni AP, Anker SD, Dahlstrom U, Filippatos G, Ponikowski P, Zannad F et al. (2013). Are hospitalized or ambulatory patients with heart failure treated in accordance with European Society of Cardiology guidelines? Evidence from 12,440 patients of the ESC Heart Failure Long-Term Registry. *Eur J Heart Fail*; 15: 1173-1184.
- Manca A, Sculpher MJ & Goeree R (2010). The analysis of multinational cost-effectiveness data for reimbursement decisions: a critical appraisal of recent methodological developments. *Pharmacoeconomics*; 28: 1079-1096.
- Marshall DA, Levy AR, Vidaillet H, Fenwick E, Slee A, Blackhouse G et al. (2004). Cost-effectiveness of rhythm versus rate control in atrial fibrillation. *Ann Intern Med*; 141: 653-661.
- McKenna C, Palmer S, Rodgers M, Chambers D, Hawkins N, Golder S et al. (2009). Cost-effectiveness of radiofrequency catheter ablation for the treatment of atrial fibrillation in the United Kingdom. *Heart*; 95: 542-549.
- McKenna C, Maund E, Sarowar M, Fox D, Stevenson M, Pepper C et al. (2009). Dronedronedrone for atrial fibrillation and atrial flutter. A Single Technology Appraisal. Centre for Reviews and Dissemination and Centre for Health Economics.
- McKenna C, Maund E, Sarowar M, Fox D, Stevenson M, Pepper C et al. (2012). Dronedronedrone for the treatment of atrial fibrillation: a NICE single technology appraisal. *Pharmacoeconomics*; 30: 35-46.
- McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Bohm M, Dickstein K et al. (2012). ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J*; 33: 1787-1847.

Mejhert M, Linder-Klingsell E, Edner M, Kahan T & Persson H (2002). Ventilatory variables are strong prognostic markers in elderly patients with heart failure. *Heart*; 88: 239-243.

Mejhert M, Kahan T, Persson H & Edner M (2006). Predicting readmissions and cardiovascular events in heart failure patients. *Int J Cardiol*; 109: 108-113.

Miller G, Randolph S, Forkner E, Smith B & Galbreath AD (2009). Long-term cost-effectiveness of disease management in systolic heart failure. *Med Decis Making*; 29: 325-333.

Mommersteeg PM, Denollet J, Spertus JA & Pedersen SS (2009). Health status as a risk factor in cardiovascular disease: a systematic review of current evidence. *Am Heart J*; 157: 208-218.

Moock J & Kohlmann T (2008). Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. *Qual Life Res*; 17: 485-495.

Mosterd A & Hoes AW (2007). Clinical epidemiology of heart failure. *Heart*; 93: 1137-1146.

Mulhern B, Bansback N, Brazier J, Buckingham K, Cairns J, Devlin N et al. (2014). Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technol Assess*; 18: vii-191.

National Institute for Health and Care Excellence (2013). *Guide to the methods of technology appraisal 2013*.

Nichol G, Kaul P, Huszti E & Bridges JF (2004). Cost-effectiveness of cardiac resynchronization therapy in patients with symptomatic heart failure. *Ann Intern Med*; 141: 343-351.

Nichols M, Townsend N, Luengo-Fernandez R, Leal J, Gray A, Scarborough P et al. (2012). European Cardiovascular Disease Statistics 2012. European Society of Cardiology.

Nieuwlaat R, Capucci A, Camm AJ, Olsson SB, Andresen D, Davies DW et al. (2005). Atrial fibrillation management: a prospective survey in ESC member countries: the Euro Heart Survey on Atrial Fibrillation. *Eur Heart J*; 26: 2422-2434.

Nieuwlaat R, Prins MH, Le Heuzey JY, Vardas PE, Aliot E, Santini M et al. (2008). Prognosis, disease progression, and treatment of atrial fibrillation patients during 1 year: follow-up of the Euro Heart Survey on atrial fibrillation. *Eur Heart J*; 29: 1181-1189.

Nilsson E (2014). Patientrapporterade utfallsmått (PROM) i kvalitetsregister för långvariga sjukdomstillstånd: Generella mått och indikatorer - utvecklingsförslag.

Nilsson J, Akerborg O, Bego-LeBagousse G, Rosenquist M & Lindgren P (2013). Cost-effectiveness analysis of dronedarone versus other anti-arrhythmic drugs for the treatment of atrial fibrillation--results for Canada, Italy, Sweden and Switzerland. *Eur J Health Econ*; 14: 481-493.

O'Connor CM, Whellan DJ, Wojdyla D, Leifer E, Clare RM, Ellis SJ et al. (2012). Factors related to morbidity and mortality in patients with chronic heart failure with systolic dysfunction: the HF-ACTION predictive risk score model. *Circ Heart Fail*; 5: 63-71.

O'Loughlin C, Murphy NF, Conlon C, O'Donovan A, Ledwidge M & McDonald K (2010). Quality of life predicts outcome in a heart failure disease management program. *Int J Cardiol*; 139: 60-67.

- Peinado R, Arribas F, Ormaetxe JM & Badia X (2010). Variation in quality of life with type of atrial fibrillation. *Rev Esp Cardiol*; 63: 1402-1409.
- Perez A, Touchette DR, DiDomenico RJ, Stamos TD & Walton SM (2011). Comparison of rate control versus rhythm control for management of atrial fibrillation in patients with coexisting heart failure: a cost-effectiveness analysis. *Pharmacotherapy*; 31: 552-565.
- Peters-Klimm F, Kunz CU, Laux G, Szecsenyi J & Muller-Tasch T (2010). Patient- and provider-related determinants of generic and specific health-related quality of life of patients with chronic systolic heart failure in primary care: a cross-sectional study. *Health Qual Life Outcomes*; 8: 98.
- Pickard AS, Tawk R & Shaw JW (2013). The effect of chronic conditions on stated preferences for health. *Eur J Health Econ*; 14: 697-702.
- Pocock SJ, Ariti CA, McMurray JJ, Maggioni A, Kober L, Squire IB et al. (2013). Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J*; 34: 1404-1413.
- Politi MC, Clayman ML, Fagerlin A, Studts JL & Montori V (2013). Insights from a conference on implementing comparative effectiveness research through shared decision-making. *J Comp Eff Res*; 2: 23-32.
- Powell J (1984). Least absolute deviations for the censored regression model. *J Econometrics*; 25: 303-325.
- Pressler SJ, Eckert GJ, Morrison GC, Murray MD & Oldridge NB (2011). Evaluation of the Health Utilities Index Mark-3 in heart failure. *J Card Fail*; 17: 143-150.
- PROM center (2014). *De nationella kvalitetsregistren*. Available at: <http://www.promcenter.se/sv/prom-i-kvalitetsregister>. Accessed 2014-10-01.
- Pullenayegum EM, Tarride JE, Xie F, Goeree R, Gerstein HC & O'Reilly D (2010). Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? *Value Health*; 13: 487-494.
- Pullenayegum EM, Wong HS & Childs A (2013). Generalized additive models for the analysis of EQ-5D utility data. *Med Decis Making*; 33: 244-251.
- Quaglini S, Shahar Y, Peleg M, Miksch S, Napolitano C, Rigla M et al. (2013). Supporting shared decision making within the MobiGuide project. *AMIA Annu Symp Proc*; 2013: 1175-1184.
- Raikou M & McGuire A (2004). Estimating medical care costs under conditions of censoring. *J Health Econ*; 23: 443-470.
- Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A et al. (2005). Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health*; 8: 521-533.
- Reynolds MR, Lavelle T, Essebag V, Cohen DJ & Zimetbaum P (2006). Influence of age, sex, and atrial fibrillation recurrence on quality of life outcomes in a population of patients with new-onset atrial fibrillation: the Fibrillation Registry Assessing Costs, Therapies, Adverse events and Lifestyle (FRACTAL) study. *Am Heart J*; 152.
- Reynolds MR, Zimetbaum P, Josephson ME, Ellis E, Danilov T & Cohen DJ (2009). Cost-effectiveness of radiofrequency catheter ablation compared with antiarrhythmic drug therapy for paroxysmal atrial fibrillation. *Circ Arrhythm Electrophysiol*; 2: 362-369.

- Reynolds MR, Morais E & Zimetbaum P (2010). Impact of hospitalization on health-related quality of life in atrial fibrillation patients in Canada and the United States: results from an observational registry. *Am Heart J*; 160: 752-758.
- Reynolds MR, Nilsson J, Akerborg O, Jhaveri M & Lindgren P (2013). Cost-effectiveness of dronedarone and standard of care compared with standard of care alone: US results of an ATHENA lifetime model. *Clinicoecon Outcomes Res*; 5: 19-28.
- Richardson J, Khan MA, Iezzi A & Maxwell A (2014). Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments. *Med Decis Making*.
- Rienstra M, McManus DD & Benjamin EJ (2012). Novel risk factors for atrial fibrillation: useful for risk prediction and clinical decision making? *Circulation*; 125: e941-e946.
- Ringborg A, Nieuwlaat R, Lindgren P, Jonsson B, Fidan D, Maggioni AP et al. (2008). Costs of atrial fibrillation in five European countries: results from the Euro Heart Survey on atrial fibrillation. *Europace*; 10: 403-411.
- Roalfe AK, Bryant TL, Davies MH, Hackett TG, Saba S, Fletcher K et al. (2012). A cross-sectional study of quality of life in an elderly population (75 years and over) with atrial fibrillation: secondary analysis of data from the Birmingham Atrial Fibrillation Treatment of the Aged study. *Europace*; 14: 1420-1427.
- Roger VL (2013). Epidemiology of heart failure. *Circ Res*; 113: 646-659.
- Ryen L & Svensson M (2014). The willingness to pay for a QALY: A review of the empirical literature. Karlstad University Working Paper in Economics.
- Salomon JA, Vos T, Hogan DR, Gagnon M, Naghavi M, Mokdad A et al. (2012). Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet*; 380: 2129-2143.
- Sartipy U, Dahlstrom U, Edner M & Lund LH (2014). Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51,043 patients from the Swedish Heart Failure Registry. *Eur J Heart Fail*; 16: 173-179.
- Schmittiel JA, Uratsu CS, Karter AJ, Heisler M, Subramanian U, Mangione CM et al. (2008). Why don't diabetes patients achieve recommended risk factor targets? Poor adherence versus lack of treatment intensification. *J Gen Intern Med*; 23: 588-594.
- Schron E, Friedmann E & Thomas SA (2014). Does health-related quality of life predict hospitalization or mortality in patients with atrial fibrillation? *J Cardiovasc Electrophysiol*; 25: 23-28.
- Siegel JE, Torrance GW, Russell LB, Luce BR, Weinstein MC & Gold MR (1997). Guidelines for pharmacoeconomic studies. Recommendations from the panel on cost effectiveness in health and medicine. Panel on cost Effectiveness in Health and Medicine. *Pharmacoeconomics*; 11: 159-168.
- Singh JA, Nelson DB, Fink HA & Nichol KL (2005). Health-related quality of life predicts future health care utilization and mortality in veterans with self-reported physician-diagnosed arthritis: the veterans arthritis quality of life study. *Semin Arthritis Rheum*; 34: 755-765.
- Sorensen J, Linde L, Ostergaard M & Hetland ML (2012). Quality-adjusted life expectancies in patients with rheumatoid arthritis--comparison of index scores from EQ-5D, 15D, and SF-6D. *Value Health*; 15: 334-339.

Soto GE, Jones P, Weintraub WS, Krumholz HM & Spertus JA (2004). Prognostic value of health status in patients with heart failure after acute myocardial infarction. *Circulation*; 110: 546-551.

Spertus J, Dorian P, Bubien R, Lewis S, Godejohn D, Reynolds MR et al. (2011). Development and validation of the Atrial Fibrillation Effect on Quality-of-Life (AFEQT) Questionnaire in patients with atrial fibrillation. *Circ Arrhythm Electrophysiol*; 4: 15-25.

Sprangers MA & Schwartz CE (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med*; 48: 1507-1515.

Steel JL, Geller DA, Robinson TL, Savkova AY, Brower DS, Marsh JW et al. (2014). Health-related quality of life as a prognostic factor in patients with advanced cancer. *Cancer*; 120: 3717-3721.

Steg PG, Alam S, Chiang CE, Gamra H, Goethals M, Inoue H et al. (2012). Symptoms, functional status and quality of life in patients with controlled and uncontrolled atrial fibrillation: data from the RealiseAF cross-sectional international registry. *Heart*; 98: 195-201.

Stull DE, Clough LA & Van DD (2001). Self-report quality of life as a predictor of hospitalization for patients with LV dysfunction: a life course approach. *Res Nurs Health*; 24: 460-469.

Subramanian U, Eckert G, Yeung A & Tierney WM (2007). A single health status question had important prognostic value among outpatients with chronic heart failure. *J Clin Epidemiol*; 60: 803-811.

Sullivan MD, Levy WC, Russo JE, Crane B & Spertus JA (2007). Summary health status measures in advanced heart failure: relationship to clinical variables and outcome. *J Card Fail*; 13: 560-568.

Tandvårds- och Läkemedelsförmånsverket (2003). *General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2)*.

Tandvårds- och Läkemedelsförmånsverket (2011). *Beslut: Multaq*.

Tidermark J & Bergstrom G (2007). Responsiveness of the EuroQol (EQ-5D) and the Nottingham Health Profile (NHP) in elderly patients with femoral neck fractures. *Qual Life Res*; 16: 321-330.

Timbie JW, Shahian DM, Newhouse JP, Rosenthal MB & Normand SL (2009). Composite measures for hospital quality using quality-adjusted life years. *Stat Med*; 28: 1238-1254.

Torrance GW (1987). Utility approach to measuring health-related quality of life. *J Chronic Dis*; 40: 593-603.

Tsuchiya A & Dolan P (2005). The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Med Decis Making*; 25: 460-467.

Van Gelder IC, Hagens VE, Bosker HA, Kingma JH, Kamp O, Kingma T et al. (2002). A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. *N Engl J Med*; 347: 1834-1840.

Vasko P, Jonsson A & Dahlstrom U (2013). Årsrapport RiksSvikt: 2012 års resultat. RiksSvikt.

Ware JE, Jr. & Sherbourne CD (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*; 30: 473-483.



Weinstein MC, Torrance G & McGuire A (2009). QALYs: the basics. *Value Health*; 12 Suppl 1: S5-S9.

Wilson IB & Cleary PD (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*; 273: 59-65.

Wolf PA, Abbott RD & Kannel WB (1991). Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*; 22: 983-988.

World Health Organization (2011). *Global atlas on cardiovascular disease prevention and control*.

World Health Organization (2013). *WHO methods and data sources for global burden of disease estimates 2000-2011*. Geneva.

Wyse DG, Waldo AL, DiMarco JP, Domanski MJ, Rosenberg Y, Schron EB et al. (2002). A comparison of rate control and rhythm control in patients with atrial fibrillation. *N Engl J Med*; 347: 1825-1833.

Zarrinkoub R, Wettermark B, Wandell P, Mejhert M, Szulkin R, Ljunggren G et al. (2013). The epidemiology of heart failure, based on data for 2.1 million inhabitants in Sweden. *Eur J Heart Fail*; 15: 995-1002.

Zethraeus N, Johannesson M, Jonsson B, Lothgren M & Tambour M (2003). Advantages of using the net-benefit approach for analysing uncertainty in economic evaluation studies. *Pharmacoeconomics*; 21: 39-48.

Zuluaga MC, Guallar-Castillon P, Lopez-Garcia E, Banegas JR, Conde-Herrera M, Olcoz-Chiva M et al. (2010). Generic and disease-specific quality of life as a predictor of long-term mortality in heart failure. *Eur J Heart Fail*; 12: 1372-1378.